

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: September 10, 2003, 14:09:48 ; Search time 34 Seconds  
(without alignments)  
478.156 Million cell updates/sec

Title: US-09-822-540a-1

Perfect score: 346  
Sequence: 1 ADRAAVPIVNLKDELLFPGW.....DVLRHPTWPKSVWHSQDPN 63

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	187	4 O8WX15	O8WX15 homo sapien
2	346	100.0	261	4 O8NG19	O8NG19 homo sapien
3	346	100.0	816	4 O8N4S4	O8N4S4 homo sapien
4	298	86.1	102	4 O96T70	O96T70 homo sapien
5	294	85.0	160	11 O9CRT2	O9CRT2 mus musculi
6	294	85.0	184	11 O9JK63	O9JK63 mus musculi
7	294	85.0	1140	11 O61434	O61434 mus musculi
8	294	85.0	1774	11 O62001	O62001 mus musculi
9	284	82.1	171	11 O9WTM5	O9WTM5 rattus norv
10	284	82.1	226	11 O9Q2D2	O9Q2D2 rattus norv
11	255	73.7	1344	13 O93419	O93419 gallus gall
12	250	72.3	1315	13 O8QHL9	O8QHL9 xenopus lae
13	233	67.3	1307	13 O8JUF7	O8JUF7 xenopus lae
14	219	63.3	1361	13 O8AWC6	O8AWC6 brachydanio
15	197	56.9	1367	11 O35206	O35206 mus musculi
16	196	56.6	1367	11 O9EOD9	O9EOD9 mus musculi

17	193	55.8	1388	4 O9Y4W4	O9Y4W4 homo sapien
18	169.5	49.0	299	5 O8WSE3	O8WSE3 drosophila
19	169.5	49.0	792	5 O8WT89	O8WT89 drosophila
20	140	40.5	650	5 O17866	O17866 caenorhabdi
21	140	40.5	778	5 O9U9K5	O9U9K5 caenorhabdi
22	140	40.5	1117	5 O9U9K7	O9U9K7 caenorhabdi
23	140	40.5	1154	5 O810G6	O810G6 caenorhabdi
24	70.5	20.4	317	4 O43467	O43467 homo sapien
25	67.5	19.5	614	4 O03989	O03989 homo sapien
26	65	18.8	958	5 O8SX15	O8SX15 drosophila
27	64	18.5	644	17 O9YAI3	O9YAI3 aeropyrum
28	63.5	18.4	266	10 O8M159	O8M159 brassica oi
29	63.5	18.4	847	16 O8Y0C4	O8Y0C4 ralbionia s
30	63	18.2	336	2 O9ZG16	O9ZG16 streptomyce
31	62.5	18.1	197	17 O8ZTQ1	O8ZTQ1 pyrobaculum
32	62.5	18.1	467	5 O9XX44	O9XX44 caenorhabdi
33	62.5	18.1	620	16 O11526	O11526 bacillus su
34	62	17.9	102	4 O8WT95	O8WT95 homo sapien
35	62	17.9	126	17 O8TU72	O8TU72 methanosarc
36	62	17.9	325	4 O9Y247	O9Y247 homo sapien
37	62	17.9	419	5 P91096	P91096 caenorhabdi
38	61.5	17.8	362	7 O9MXH0	O9MXH0 pan troglod
39	61.5	17.8	1172	12 O9E1Y8	O9E1Y8 cercopithec
40	61	17.6	738	11 O9Z213	O9Z213 rattus norv
41	60.5	17.5	236	3 O8X1J5	O8X1J5 cryptococcu
42	60.5	17.5	862	10 O9FT17	O9FT17 lycopersico
43	60	17.3	258	17 P94920	P94920 methanosarc
44	60	17.3	258	17 O9C4S2	O9C4S2 methanosarc
45	60	17.3	260	17 O8Q0G2	O8Q0G2 methanosarc

## ALIGNMENTS

### RESULT 1

O8WX15  
AC O8WX15  
DT 01-MAR-2002 (TRENBLREL, 20, Created)  
DT 01-MAR-2002 (TRENBLREL, 20, Last sequence update)  
DT 01-MAR-2002 (TRENBLREL, 20, Last annotation update)  
DE Collagen XVIII (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21409408; PubMed=11517600;  
RA Feng Y., Cui L.B., Liu C.X., Ma Q.J.  
RT "Inhibition effect in vitro of purified endostatin expressed in Pichia  
RT pastoris".  
RL Sheng Wu Gong Cheng Xue Bao 17:278-282 (2001).  
DR EMBL; AF416592; AAL37720.1;  
FT  
SQ SEQUENCE 187 AA; 20448 MW; 72B1047D85838CD3 CRC64;

Query Match 100.0%; Score 346; DB 4; Length 187;  
Best Local Similarity 100.0%; Pred. No. 3e-34; 0; Indels 0; Gaps 0;  
Matches 63; Conservative 0; Mismatches 0;

OY 1 ADRAAVPIVNLKDELLFPGWALFSGSGEPLKRGARIFSPDGKDVLRHPTWPKSVWHS 60  
DB 68 ADRAAVPIVNLKDELLFPGWALFSGSGEPLKRGARIFSPDGKDVLRHPTWPKSVWHS 127

OY 61 DPN 63  
DB 128 DPN 130

RESULT 2  
O8NG19  
ID O8NG19 PRELIMINARY; PRT; 261 AA.

AC Q8NG19;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Multi-Functional Protein MFP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF282883; AAM52249.1; -  
 SO SEQUENCE 261 AA; 26745 MW; CA60C920AF3E90E5 CRC64;  
 Query Match 100.0%; Score 346; DB 4; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-34;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ADRAAVPIVNLKDELLFPGSEGLPKRGARIFSGDKDVLRHPTWPKSVWHS 60  
 DB 142 ADRAAVPIVNLKDELLFPGSEGLPKRGARIFSGDKDVLRHPTWPKSVWHS 201  
 QY 61 DPN 63  
 DB 202 DPN 204  
 RESULT 3  
 Q8N4S4 PRELIMINARY; PRT; 816 AA.  
 ID Q8N4S4  
 AC Q8N4S4;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Similar to collagen, type XVIII, alpha 1 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Renal adenocarcinoma;  
 RA Strausberg R.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC033715; AAH33715.1; -  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 6.  
 DR ProDom: PDD00007; Collagen; 1.  
 KW Collagen.  
 FT NON TER  
 SQ SEQUENCE 816 AA; 82553 MW; 5D539B2946694F86 CRC64;  
 Query Match 100.0%; Score 346; DB 4; Length 816;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-33;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ADRAAVPIVNLKDELLFPGSEGLPKRGARIFSGDKDVLRHPTWPKSVWHS 60  
 DB 697 ADRAAVPIVNLKDELLFPGSEGLPKRGARIFSGDKDVLRHPTWPKSVWHS 756  
 QY 61 DPN 63  
 DB 757 DPN 759  
 RESULT 4  
 Q86T70 PRELIMINARY; PRT; 102 AA.  
 ID Q86T70  
 AC Q86T70;  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE Endostatin variant (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Deininger M.H., Trautmann K., Schluessener H.J.;  
 RT "Endostatin promotes delayed secondary damage following traumatic brain injury."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF333247; AA050626.1; -  
 FT NON TER  
 SQ SEQUENCE 102 AA; 11147 MW; ECAC47AA6420947D CRC64;  
 Query Match 86.1%; Score 298; DB 4; Length 102;  
 Best Local Similarity 98.2%; Pred. No. 1.1e-28;  
 Matches 55; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 8 IYNLDELLFPGSEGLPKRGARIFSGDKDVLRHPTWPKSVWHS 63  
 DB 1 IYNLDELLFPGSEGLPKRGARIFSGDKDVLRHPTWPKSVWHS 66  
 RESULT 5  
 Q9CRT2 PRELIMINARY; PRT; 160 AA.  
 ID Q9CRT2  
 AC Q9CRT2;  
 DT 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE Procollagen, type XVIII, alpha 1 (Fragment).  
 GN COL18A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; Pubmed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Strubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Botteilli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereis P.,  
 RA Nordone P., Ring B., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Sato K., Schenck C., Wang K.H., Weitz C., Whitaker C., Wilting L.,  
 RA Wyshaw-Borje A., Yoshida K., Haegawa Y., Kawaji H., Kohlschki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK014292; BAB29249.1; -  
 DR HSP; P39061; IKOE.  
 DR MGI; MGI:88451; Coll18a1.  
 FT NON TER  
 SQ SEQUENCE 160 AA; 17725 MW; 60F853D777C375D2 CRC64;  
 Query Match 85.0%; Score 294; DB 11; Length 160;  
 Best Local Similarity 81.0%; Pred. No. 5.5e-28;  
 Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 ADRAVPVIVNLKDELFPSEWALFSGSEGLPKGARIFSPDGKDVLRHPTWPKSVWHS 60
DB 40 ADRCVPIVNLKDEVLPSPWDLFSGSQOLPGARIFSPDGRDVLRRHPAMPQKSVWHS 99
QY 61 DPN 63
DB 100 DPS 102

RESULT 6
QYK63 PRELIMINARY; PRT; 184 AA.
AC Q9UK63:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Endostatin (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Chinese Kunming;
RA Jia S., Zhu F., King G., Yu Y., Duan C., Xiu R.-J., He F.;
RT "Anticancer treatment of targeted fusion protein delivery to tumor
RT neovasculature."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF257775; AAF69009.1; -
DR HSSP; P39061; 1KOE.
FT NON_TER 1
FT NON_TER 184
SQ SEQUENCE 184 AA; 20376 MW; AC06F9D8D103412A CRC64;

Query Match
Best Local Similarity 85.0%; Score 294; DB 11; Length 184;
Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAVPVIVNLKDELFPSEWALFSGSEGLPKGARIFSPDGKDVLRHPTWPKSVWHS 60
DB 64 ADRCVPIVNLKDEVLPSPWDLFSGSQOLPGARIFSPDGRDVLRRHPAMPQKSVWHS 123
QY 61 DPN 63
DB 124 DPS 126

RESULT 7
QYK63 PRELIMINARY; PRT; 1140 AA.
AC Q61434:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Collagen (Fragment).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC ABe N., Muragaki Y., Yoshioke H., Inoue H., Ninomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region."
RL Cell. Mol. Biol. Res. 196; 576-582 (1993).
DR EMBL; D17546; BAA04483.1; -
DR HSSP; P39061; 1KOE.
MGD; MGI:88449; Col15a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF02191; Collagen; 8.
DR Pfam; PF02210; TSPN; 1.

```

```

DR Prodom; PD000007; Collagen; 1.
KW Collagen.
FT NON_TER 1
SQ SEQUENCE 1140 AA; 115156 MW; 8B0C7E6862B3BDFE CRC64;

Query Match
Best Local Similarity 85.0%; Score 294; DB 11; Length 1140;
Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAVPVIVNLKDELFPSEWALFSGSEGLPKGARIFSPDGKDVLRHPTWPKSVWHS 60
DB 1020 ADRCVPIVNLKDEVLPSPWDLFSGSQOLPGARIFSPDGRDVLRRHPAMPQKSVWHS 1079
QY 61 DPN 63
DB 1080 DPS 1082

RESULT 8
QYK63 PRELIMINARY; PRT; 1774 AA.
AC Q62001: Q60672;
ID Q62001; Q60672;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Procollagen, type XVIII, alpha 1 precursor (XVIII) collagen
DE (Procollagen, type XVIII, alpha 1) (Alpha-1 type XVIII collagen).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PRIMARY TAIL CULTURE;
RX MEDLINE=94245707; PubMed=8188673;
RA Rehn M., Hinclicka E., Philajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain."
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M., Philajaniemi T.;
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen."
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [3]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=95181468; PubMed=7876242;
RA Rehn M., Philajaniemi T.;
RT "Identification of three N-terminal ends of type XVIII collagen chains
RT and tissue-specific differences in the expression of the corresponding
RT transcripts. The longest form contains a novel motif homologous to rat
RT and Drosophila frizzled proteins."
RL J. Biol. Chem. 270:4705-4711(1995).
DR EMBL; U03715; AAC52903.1; -
DR EMBL; U03716; AAC52903.1; JOINED.
DR EMBL; U03718; AAC52903.1; JOINED.
DR EMBL; U34607; AAC52903.1; JOINED.
DR EMBL; U34608; AAC52903.1; JOINED.
DR EMBL; U34609; AAC52903.1; JOINED.
DR EMBL; U34610; AAC52903.1; JOINED.
DR EMBL; U34611; AAC52903.1; JOINED.
DR EMBL; U34612; AAC52903.1; JOINED.
DR EMBL; U34613; AAC52903.1; JOINED.
DR HSSP; U1637; AAC52179.1; -
DR HSSP; P39061; 1KOE.
MGD; MGI:88451; Col18a1.
DR InterPro; IPR000087; Collagen.

```

DR InterPro; IPR000024; Fz domain.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF01391; Collagen; 8.  
 DR Pfam; PF01392; Fz; 1.  
 DR Pfam; PF02210; TSPN; 1.  
 DR ProDom; PD000007; Collagen; 1.  
 DR SMART; SM00063; FRI; 1.  
 DR SMART; SM00282; LamG; 1.  
 DR SMART; SM00210; TSPN; 1.  
 DR PROSITE; PS00038; FZ; 1.  
 KW Collagen; Signal.  
 SQ SEQUENCE 1774 AA; 182229 MW; CF4D9C9E88EF232 CRC64;

Query Match 85.0%; Score 294; DB 11; Length 1774;  
 Best Local Similarity 81.0%; Pred. No. 8.3e-27;  
 Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSEWALFSGSEGLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
 DB 1654 ADRAAVPIVNLKDELFPSEWALFSGSQGLQGPARIFFPDGRDVLRRHAWPKSVWHS 1713  
 QY 61 DPN 63  
 DB 1714 DPS 1716

## RESULT 9

Q9WDM5 PRELIMINARY; PRT; 171 AA.  
 AC Q9WDM5; 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Collagen type XVIII, alpha (1) chain (Fragment).  
 GN COL18A1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Jia J.D., Bauer M., Eberpaecher U., Donner P., Schuppan D.;  
 RT "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Jia J.D., Bauer M., Sedlacek N., Ruenl M., Riecken E.O., Schuppan D.;  
 RT "Temporospatial expression of collagen XVIII/endostatin in acute and  
 chronic liver injuries";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ236873; CAB44263.1; -.  
 DR HSSP; P39061; IKOE.  
 FT NON\_TER 1  
 FT NON\_TER 171  
 SQ SEQUENCE 171 AA; 18933 MW; 81BE2E3FC0C8E72 CRC64;

Query Match 82.1%; Score 284; DB 11; Length 171;  
 Best Local Similarity 79.4%; Pred. No. 9.8e-27;  
 Matches 50; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSEWALFSGSEGLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
 DB 56 ADRAAVPIVNLKDELFPSEWALFSGSQGLQGPARIFFPDGRDVLRRHAWPKSVWHS 115  
 QY 61 DPN 63  
 DB 116 DPS 118

## RESULT 10

-

Q9QZD2 PRELIMINARY; PRT; 226 AA.  
 AC Q9QZD2; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE Collagen XVIII (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=20227226; PubMed=10766159;  
 RA Perletti G., Concarl P., Giardini R., Marras E., Piccinini F.,  
 RA Folkmann J., Chen L.;  
 RT "Antitumor activity of endostatin against carcinogen-induced rat  
 primary mammary tumors";  
 RL Cancer Res. 60:1793-1796(2000).  
 DR EMBL; AF189709; AAF00975.1; -.  
 DR HSSP; P39061; IKOE.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 226 AA; 25350 MW; 38B83C0486C0B949 CRC64;

Query Match 82.1%; Score 284; DB 11; Length 226;  
 Best Local Similarity 79.4%; Pred. No. 1.3e-26;  
 Matches 50; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSEWALFSGSEGLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
 DB 106 ADRAAVPIVNLKDELFPSEWALFSGSQGLQGPARIFFPDGRDVLRRHAWPKSVWHS 165  
 QY 61 DPN 63  
 DB 166 DPS 168

## RESULT 11

O93419 PRELIMINARY; PRT; 1344 AA.  
 AC O93419; 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 23, Last annotation update)  
 DE Collagen XVIII precursor.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98411346; PubMed=9738008;  
 RA Halfter W., Dong S., Schurer B., Cole G.J.;  
 RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan";  
 RL J. Biol. Chem. 273:25404-25412(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Halfter W., Dong S.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF083440; AAC33294.2; -.  
 DR HSSP; P39061; IKOE.

DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF01391; Collagen; 8.  
 DR Pfam; PF02210; TSPN; 1.  
 DR ProDom; PD000007; Collagen; 1.  
 DR SMART; SM00282; LamG; 2.  
 DR SMART; SM00210; TSPN; 1.  
 KW Collagen; Signal.  
 FT SIGNAL 1  
 FT SIGNAL 26  
 \* POTENTIAL.

SQ SEQUENCE 1344 AA; 137402 MW; 7AA366E4FE940CCD CRC64;  
 Query Match 73.7%; Score 255; DB 13; Length 1344;  
 Best Local Similarity 72.1%; Pred. No. 3.5e-22;  
 Matches 44; Conservative 9; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 ADRAAVPIVNLKDELLFPSWEALFSGSEGLKPGARIFSGDKDVLRHPTWPKSWHGS 60  
 DB 1224 ADRTAVPIVNLKDEVLFSNWELFTGSEAPLRAGABILSPDGRDIIQDSAMWPKSIWHGS 1283  
 QY 61 D 61  
 DB 1284 D 1284  
 RESULT 12  
 Q8QHL9 PRELIMINARY; PRT; 1315 AA.  
 ID Q8QHL9;  
 AC 08QHL9;  
 DT 01-JUN-2002 (TRENBLrel. 21, Created)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Type XVIII collagen alpha1 chain.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NCBI\_TaxID=8355;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Ichino T., Sekimizu K., Natori S., Kubo T.;  
 RT "Identification and characterization of genes expressed selectively in  
 the regenerating tail of *Xenopus laevis* tadpole."  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB047066; BAB84674.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF02210; TSPN; 1.  
 DR ProDom; PD000007; Collagen; 1.  
 DR SMART; SM00210; TSPN; 1.  
 KW Collagen.  
 SQ SEQUENCE 1315 AA; 134946 MW; 0C56C235DE058365 CRC64;  
 Query Match 72.3%; Score 250; DB 13; Length 1315;  
 Best Local Similarity 72.1%; Pred. No. 1.4e-21;  
 Matches 44; Conservative 8; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 ABRRAVPIVNLKDELLFPSWEALFSGSEGLKPGARIFSGDKDVLRHPTWPKSWHGS 60  
 DB 1195 ADRGVOIVNLKDEVLFSNWELFTGSEAPLRAGABILSPDGRDIIQDSAMWPKSIWHGS 1254  
 QY 61 D 61  
 DB 1255 D 1255  
 RESULT 13  
 Q8JFF7 PRELIMINARY; PRT; 1307 AA.  
 ID Q8JFF7;  
 AC 08JFF7;  
 DT 01-OCT-2002 (TRENBLrel. 22, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Type XVIII collagen short variant.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NCBI\_TaxID=8355;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=22166979; PubMed=12175494;

RA Elamra H., Peterson J., Pihlajaniemi T., Destrée O.;  
 RT "Cloning of three variants of type XVIII collagen and their expression  
 patterns during *Xenopus laevis* development."  
 RL Mech. Dev. 114:109-113(2002).  
 DR EMBL; AY052763; AAL14257.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen; 6.  
 DR Pfam; PF02210; TSPN; 1.  
 DR SMART; SM00210; TSPN; 1.  
 KW Collagen.  
 SQ SEQUENCE 1307 AA; 134153 MW; D53EDBF3DE34976 CRC64;  
 Query Match 67.3%; Score 233; DB 13; Length 1307;  
 Best Local Similarity 67.2%; Pred. No. 1.6e-19;  
 Matches 41; Conservative 9; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 ADRAAVPIVNLKDELLFPSWEALFSGSEGLKPGARIFSGDKDVLRHPTWPKSWHGS 60  
 DB 1187 ADRGVOIVNLKDEVLFSNWELFTGSEAPLRAGABILSPDGRDIIQDSAMWPKSIWHGS 1246  
 QY 61 D 61  
 DB 1247 D 1247  
 RESULT 14  
 Q8AMC6 PRELIMINARY; PRT; 361 AA.  
 ID Q8AMC6;  
 AC 08AMC6;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Collagen XVIII (Fragment).  
 GN COL18A1.  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 NCBI\_TaxID=7955;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RA Hattex Z., Morvan-Dubois G., Thisee B., Garrone R., Le Guellec D.;  
 RT "Sequence and embryonic expression of collagen XVIII NC11 domain  
 (endostatin) in the zebrafish."  
 RL Submitted (Jul-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A494637; CAD38825.1; -  
 FT NON\_TER  
 SQ SEQUENCE 361 AA; 40222 MW; 3C5A0F8479D26735 CRC64;  
 Query Match 63.3%; Score 219; DB 13; Length 361;  
 Best Local Similarity 62.3%; Pred. No. 1.9e-18;  
 Matches 38; Conservative 10; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 ADRAAVPIVNLKDELLFPSWEALFSGSEGLKPGARIFSGDKDVLRHPTWPKSWHGS 60  
 DB 241 SDRRTPIVNLKDVLFSSWELFTGSEAPLRAGABILSPDGRDIIQDSAMWPKSIWHGS 300  
 QY 61 D 61  
 DB 301 D 301  
 RESULT 15  
 Q35206 PRELIMINARY; PRT; 1367 AA.  
 ID Q35206;  
 AC 035206;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Type XV collagen.  
 GN COL15A1.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97480713; PubMed=9339358;  
 RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,  
 RA Pihlajaniemi T.;  
 RT "Cloning of mouse type XV collagen sequences and mapping of the  
 RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1  
 RT (XV) collagen sequences indicates divergence in the number of small  
 RT collagenous domains.";  
 RL Genomics 45:31-41(1997).  
 DR EMBL; AF011450; AAC53387.1; -.  
 DR HSSP; P39061; IKOE.  
 DR MGD; MG1:88449; COL15a1.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF01391; Collagen; 5.  
 DR Pfam; PF02210; TSPN; 1.  
 DR ProDom; PD000007; Collagen; 1.  
 DR SMART; SM00282; LamG; 1.  
 DR SMART; SM00210; TSPN; 1.  
 KW Collagen.  
 SO SEQUENCE 1367 AA; 140525 MW; A483A1254AF3A8EC CRC64;

Query Match 56.9%; Score 197; DB 11; Length 1367;

Best Local Similarity 52.4%; Pred. No. 4.2e-15;  
 Matches 33; Conservative 14; Mismatches 16; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
 DB 1247 AERFGPIVNLKQVLFNNWDSIFSGGQGPNTHIPYSPGRDVMTPSPQKVWHS 1306  
 QY 61 DPN 63  
 DB 1307 NPH 1309

Search completed: September 10, 2003, 14:11:14  
 Job time : 37 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 14:09:48 ; Search time 11 Seconds

(without alignments)  
269.335 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346

Sequence: 1 ADRAAVPIVNLKDELLPSPW.....DYLRRPTWPKSVHGSDPN 63

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt 41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	346	100.0	1516	1 CAIH_HUMAN	P39060 homo sapien
2	294	85.0	1527	1 CAIH_MOUSE	P39061 mus musculu
3	193	55.8	1368	1 CAIE_HUMAN	P39059 homo sapien
4	63.5	18.4	266	1 CB21_SINAI	P13851 sinapis alb
5	61	17.6	683	1 AC31_KLULA	O60031 kluyveromyc
6	61	17.6	1045	1 PR75_SRRMA	P09489 serratia ma
7	60	17.3	694	1 TKT1_RHIME	P58333 rhizobium m
8	58.5	16.9	1184	1 ALAC_ARATH	P57792 arabidopsis
9	58.5	16.9	1720	1 FTSH_CHLVU	P56369 chlorella v
10	58	16.8	662	1 AC3A_SCHPO	P78773 schizosacch
11	57.5	16.6	553	1 FXCI_HUMAN	O12948 homo sapien
12	57	16.5	683	1 KPCL_MOUSE	P23298 mus musculu
13	56.5	16.3	770	1 TRPG_ASPNG	P05328 aspergillus
14	56.5	16.3	884	1 YP67_MYCTU	O50654 mycobacteri
15	56	16.2	651	1 SYM_METUA	O58659 methanococc
16	55.5	16.0	727	1 NUAM_HUMAN	P28331 homo sapien
17	55.5	16.0	1174	1 KCRF_STRPU	P18294 strongyloce
18	55	15.9	339	1 XAP5_HUMAN	O14330 homo sapien
19	55	15.9	380	1 OES4_DROES	P29404 drosophila
20	55	15.9	440	1 YAJ3_SCHPO	O09710 schizosacch
21	55	15.9	444	1 PAPA_CANFA	O28622 c platelatec
22	55	15.9	545	1 AIRE_HUMAN	O43918 homo sapien
23	55	15.9	673	1 XRP2_PIG	O95333 sus scrofa
24	55	15.9	683	1 KPCL_RAT	O64617 rattus norv
25	54.5	15.8	1861	1 MAP2_RAT	P15146 rattus norv
26	54	15.6	215	1 COAT_PMY	P15596 papaya mosa
27	54	15.6	434	1 KP58_MOUSE	P24788 mus musculu
28	54	15.6	436	1 KP58_RAT	P46892 rattus norv
29	54	15.6	494	1 VABR_PLAFA	O25691 plasmodium
30	54	15.6	682	1 KPCL_HUMAN	P43723 homo sapien
31	54	15.6	983	1 YOG4_CAREL	O09499 caenorhabdi
32	53.5	15.5	310	1 Y160_BUCAP	O8K934 buchera ap
33	53.5	15.5	394	1 Y129_CHLUP	O9JX22 chlamydia p

34	53.5	15.5	609	1 SYD_CAUCR	O94734 caulobacter
35	53.5	15.5	705	1 Y450_RHISN	P56566 rhizobium s
36	53	15.3	436	1 KP58_HUMAN	P21127 homo sapien
37	53	15.3	549	1 SYR_ARCFU	O29368 archaeoglob
38	53	15.3	695	1 TKT2_RHIME	P56900 rhizobium m
39	52.5	15.2	74	1 SOM1_YEAST	O05676 saccharomyc
40	52.5	15.2	282	1 HKCC_HUMAN	P31275 homo sapien
41	52.5	15.2	317	1 GH_RAT	O62867 rattus norv
42	52.5	15.2	350	1 FLNA_TREPA	P18193 treponema p
43	52.5	15.2	363	1 DCAM_SPIOL	P46255 spinacia ol
44	52.5	15.2	366	1 YBAB_BACCI	P48843 bacillus ci
45	52.5	15.2	372	1 OPBD_MOUSE	P32300 mus musculu

## ALIGNMENTS

RESULT 1  
ID CAIH\_HUMAN STANDARD; PRT; 1516 AA.  
AC P39060; Q9UK38; Q9Y6Q7; Q9Y6Q8;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(XVII) chain precursor (Contains: Endostatin).  
GN COL18A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98164096; PubMed=9503365;  
RA Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.,  
RT "Complete primary structure of two variant forms of human type XVIII  
RT collagen and tissue-specific differences in the expression of the  
RT corresponding transcripts."  
RL Matrix Biol. 16:319-328(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20289799; PubMed=10830953;  
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,  
RA Soeda E., Onki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,  
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordiek G., Hornischer K., Brandt P.,  
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecher H.,  
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dageand E.,  
RA Mehrezyer S., Borzym K., Gardiner K., Nisetic D., Francis F.,  
RA Lehtech H., Reinhardt R., Yaspo M.-L.,  
RT "The DNA sequence of human chromosome 21."  
RL Nature 405:311-319(2000).  
RN [3]  
RP SEQUENCE OF 834-1516 FROM N.A.  
RX MEDLINE=94245237; PubMed=8188291;  
RA Oh S.P., Warman M.L., Seidlin M.F., Cheng S., Knoll J.H., Timmons S.,  
RA Olsen B.R.,  
RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen  
RT and localization of the alpha 1(XVII) collagen gene to mouse  
RT chromosome 10 and human chromosome 21."  
RL Genomics 19:494-499(1994).  
RN [4]  
RP SEQUENCE OF 1334-1516 FROM N.A.  
RX TISSUE=Placenta;  
RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.,  
RT "Cloning and expression of human endostatin gene in Escherichia  
RT coli."  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP INVOLVEMENT IN KNOBLOCH SYNDROME.

	DR	SMART; SMO0282; Lamg; 1.	
	KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;	
	KM	Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;	
	KW	Polyomphlem; 3D-structure.	
	FT SIGNAL	1 23 POTENTIAL.	
	FT CHAIN	24 1516 COLLAGEN ALPHA 1(XVII) CHAIN.	
	FT DOMAIN	1334 1516 ENDOSTATIN.	
	FT DOMAIN	24 516 NONHELICAL REGION 1 (NC1).	
	FT DOMAIN	517 550 TRIPLE-HELICAL REGION 1 (COL1).	
	FT DOMAIN	551 560 NONHELICAL REGION 2 (NC2).	
	FT DOMAIN	561 640 TRIPLE-HELICAL REGION 2 (COL2).	
	FT DOMAIN	* 641 664 NONHELICAL REGION 3 (NC3).	
	FT DOMAIN	* "TRIPLE-HELICAL REGION 3 (COL3).	
	FT DOMAIN	NONHELICAL REGION 4 (NC4).	
	FT DOMAIN	787 809 TRIPLE-HELICAL REGION 4 (COL4).	
	FT DOMAIN	810 892 NONHELICAL REGION 5 (NC5).	
	FT DOMAIN	893 906 TRIPLE-HELICAL REGION 5 (COL5).	
	FT DOMAIN	907 948 NONHELICAL REGION 6 (NC6).	
	FT DOMAIN	949 961 TRIPLE-HELICAL REGION 6 (COL6).	
	FT DOMAIN	962 1034 TRIPLE-HELICAL REGION 7 (NC7).	
	FT DOMAIN	1035 1044 NONHELICAL REGION 7 (COL7).	
	FT DOMAIN	1045 1077 TRIPLE-HELICAL REGION 8 (NC8).	
	FT DOMAIN	1078 1089 TRIPLE-HELICAL REGION 8 (COL8).	
	FT DOMAIN	1090 1111 NONHELICAL REGION 9 (NC9).	
	FT DOMAIN	1112 1118 TRIPLE-HELICAL REGION 9 (COL9).	
	FT DOMAIN	1119 1173 NONHELICAL REGION 10 (NC10).	
	FT DOMAIN	1174 1186 TRIPLE-HELICAL REGION 10 (COL10).	
	FT DOMAIN	1187 1204 NONHELICAL REGION 11 (NC11).	
	FT DOMAIN	1205 1516 N-LINKED (GLCNAc . . . ) (POTENTIAL).	
	FT CARBOHYD	68 N-LINKED (GLCNAc . . . ) (POTENTIAL).	
	FT CARBOHYD	129 N-LINKED (GLCNAc . . . ) (POTENTIAL).	
	FT CARBOHYD	164 N-LINKED (GLCNAc . . . ) (POTENTIAL).	
	FT CARBOHYD	691 O-LINKER (GALNAC . . . ) (POTENTIAL).	
	FT CARBOHYD	1329 /FtId=CARB_000150.	
	FT DISULFD	BY SIMILARITY.	
	FT DISULFD	1366 CELL ATTACHMENT SITE (POTENTIAL).	
	FT SITE	1468 CELLS AT TISSUE SITE (POTENTIAL).	
	VARSPLIC	1095 Missing (in isoform short).	
	FtId=VSP_001155.		
	VARSPLIC	1 180 HTTEGTLPACTPSPTSGRPAPITGSVPSSS -> MA PCRWPMRRRLIDVLAVLVLLGVRAASAEK (in isoform short).	
	FT VARIANT	1437 /FtId=VSP_001156. D -> N (increased risk of developing prostate cancer). . /FtId=VAR_012709. E -> S (IN REF. 2). I -> V (IN REF. 3). V -> L (IN REF. 3). P -> R (IN REF. 3). R -> P (IN REF. 3). P -> L (IN REF. 3). P -> L (IN REF. 3). A -> P (IN REF. 3). L -> K (IN REF. 3). P -> A (IN REF. 3). P -> A (IN REF. 3). P -> PGPG (IN REF. 2). G -> GO (IN REF. 3). R -> G (IN REF. 3). LR -> CG (IN REF. 3). R -> T (IN REF. 4). S -> Y (IN REF. 4). SC SEQUENCE 1516 AA; 153840 MW; 3C7OF29PA4476EE76 CRC64;	
	Query Match	Best Local Similarity	Score 346; DB 1; Length 1516; Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



DB 1397 ADRAAPVIVUKDELLFPSSWEALFSGSEGPLKFCGARIFFDQKDVLRHPTWPKQSVWHS 1456  
 OY 61 DPN 63  
 DB 1457 DPN 1459

RESULT 2  
 CAH\_MOUSE STANDARD; PRT; 1527 AA.  
 ID P39061; 061437; 062002;  
 AC 01-FEB-1995 (Rel. 31, Created)  
 DT 15-SEP-2003 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DB Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].  
 GN COL18A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RC STRAIN=BALB/C; TISSUE=Liver;  
 RX MEDLINE=94245707; PubMed=818673;  
 RA Rehn M.V., Hinkka E., Pihlajaniemi T.;  
 RT "Primary structure of the alpha 1 chain of mouse type XVII collagen, partial structure of the corresponding gene, and comparison of the alpha 1(XVII) chain with its homologue, the alpha 1(XV) collagen chain."  
 RT J. Biol. Chem. 269:13929-13935(1994).  
 RN (2)  
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
 RA Rehn M., Hinkka E., Pihlajaniemi T.;  
 RT "Characterization of the mouse gene for the alpha-1 chain of type XVII collagen (COL18A1) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters."  
 RT Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE OF 213-1140 FROM N.A. (ISOFORM SHORT).  
 RX MEDLINE=94240112; PubMed=8183894;  
 RA Rehn M.V., Pihlajaniemi T.;  
 RT "Alpha 1(XVII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen."  
 RT Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).  
 RN (4)  
 RP SEQUENCE OF 240-1527 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=94240111; PubMed=8183893;  
 RA Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;  
 RT "Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins."  
 RT Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).  
 RN (5)  
 RP CHARACTERIZATION OF ENDOSTATIN AND PARTIAL SEQUENCE.  
 RX MEDLINE=97160848; PubMed=9008168;  
 RA O'Reilly M.S., Boehm T., Sheng Y., Fukai N., Vasios G., Lane W.S., Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;  
 RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor growth."  
 RT Cell 88:277-285(1997).  
 RN (6)  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.  
 RX MEDLINE=98169382; PubMed=9501087;  
 RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;  
 RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A resolution."  
 RT EMBO J. 17:1656-1664(1998).  
 CC - FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH

CC FACTOR SIGNALING.  
 CC - ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=P39061-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P39061-2; Sequence=VSP 001157, VSP 001158;  
 CC - PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC - SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb.ch/announce/> or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC -----  
 DR EMBL; L16898; AAA37434.1; -  
 DR EMBL; U03714; AAC20657.1; -  
 DR EMBL; U03715; AAC52901.1; JOINED.  
 DR EMBL; U34606; AAC52901.1; JOINED.  
 DR EMBL; U34608; AAC52901.1; JOINED.  
 DR EMBL; U34609; AAC52901.1; JOINED.  
 DR EMBL; U34610; AAC52901.1; JOINED.  
 DR EMBL; U34611; AAC52901.1; JOINED.  
 DR EMBL; U34612; AAC52901.1; JOINED.  
 DR EMBL; U34613; AAC52901.1; JOINED.  
 DR EMBL; U03716; AAC52901.1; JOINED.  
 DR EMBL; U03718; AAC52901.1; JOINED.  
 DR EMBL; U03715; AAC52902.1; JOINED.  
 DR EMBL; U34608; AAC52902.1; JOINED.  
 DR EMBL; U34609; AAC52902.1; JOINED.  
 DR EMBL; U34610; AAC52902.1; JOINED.  
 DR EMBL; U34611; AAC52902.1; JOINED.  
 DR EMBL; U34612; AAC52902.1; JOINED.  
 DR EMBL; U03716; AAC52902.1; JOINED.  
 DR EMBL; U03718; AAC52902.1; JOINED.  
 DR EMBL; U11636; AAC52178.1; -  
 DR EMBL; L22545; AAA19787.1; -  
 DR PIR; A56101; A56101.  
 DR PDB; 1KOE; 16-FEB-99.  
 DR PDB; 1DY0; 11-APR-00.  
 DR PDB; 1DY1; 21-JAN-01.  
 DR MGD; MGI:88451; Col18a1.  
 DR GO; GO:0001525; P:angiogenesis; IMP.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF01391; Collagen; 8.  
 DR Pfam; PF02210; TSPN; 1.  
 DR ProDom; PD000007; C1g\_helix; 1.  
 DR SMART; SM00282; Lamg; 1.  
 DR SMART; SM00210; TSPN; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;  
 KW 3D-structure.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1527  
 FT CHAIN 1344 1527  
 FT DOMAIN 27 538  
 FT DOMAIN 539 565  
 FT DOMAIN 566 575  
 FT DOMAIN 576 649  
 FT DOMAIN 650 673  
 FT DOMAIN 674 795  
 FT DOMAIN 796 818  
 FT DOMAIN 819 901  
 FT DOMAIN 902 915  
 FT POTENTIAL.  
 FT COLLAGEN ALPHA 1(XVII) CHAIN.  
 FT ENDOSTATIN.  
 FT NONHELICAL REGION 1 (NC1).  
 FT TRIPLE-HELICAL REGION 1 (COL1).  
 FT NONHELICAL REGION 2 (NC2).  
 FT TRIPLE-HELICAL REGION 2 (COL2).  
 FT NONHELICAL REGION 3 (NC3).  
 FT TRIPLE-HELICAL REGION 3 (COL3).  
 FT NONHELICAL REGION 4 (NC4).  
 FT TRIPLE-HELICAL REGION 4 (COL4).  
 FT NONHELICAL REGION 5 (NC5).



KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Glycoprotein; Signal.  
 FT SIGNAL  
 FT CHAIN 1  
 FT DOMAIN 26 1388 COLLAGEN ALPHA 1(XV) CHAIN.  
 FT DOMAIN 26 555 POTENTIAL.  
 FT DOMAIN 556 573 NONHELIICAL REGION 1 (NC1).  
 FT DOMAIN 574 618 TRIPE-HELIICAL REGION 1 (COL1).  
 FT DOMAIN 619 732 TRIPE-HELIICAL REGION 2 (NC2).  
 FT DOMAIN 733 763 NONHELIICAL REGION 3 (NC3).  
 FT DOMAIN 764 798 TRIPE-HELIICAL REGION 3 (COL3).  
 FT DOMAIN 799 822 NONHELIICAL REGION 4 (NC4).  
 FT DOMAIN 823 867 TRIPE-HELIICAL REGION 4 (COL4).  
 FT DOMAIN 868 878 NONHELIICAL REGION 5 (NC5).  
 FT DOMAIN 879 949 TRIPE-HELIICAL REGION 5 (COL5).  
 FT DOMAIN 950 983 TRIPE-HELIICAL REGION 6 (NC6).  
 FT DOMAIN 984 1013 TRIPE-HELIICAL REGION 6 (COL6).  
 FT DOMAIN 1014 1027 NONHELIICAL REGION 7 (NC7).  
 FT DOMAIN 1028 1045 TRIPE-HELIICAL REGION 7 (COL7).  
 FT DOMAIN 1046 1052 NONHELIICAL REGION 8 (NC8).  
 FT DOMAIN 1053 1107 TRIPE-HELIICAL REGION 8 (COL8).  
 FT DOMAIN 1108 1117 NONHELIICAL REGION 9 (NC9).  
 FT DOMAIN 1118 1132 TRIPE-HELIICAL REGION 9 (COL9).  
 FT DOMAIN 1133 1388 NONHELIICAL REGION 10 (NC10).  
 FT DOMAIN 358 358 4 X TANDEN REPEATS.  
 FT REPEAT 358 408 1.  
 FT REPEAT 409 459 2.  
 FT REPEAT 460 509 3.  
 FT REPEAT 510 555 4.  
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 687 687 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 10 10 C -> S (IN REF. 2).  
 FT CONFLICT 49 49 D -> V (IN REF. 2).  
 FT CONFLICT 95 95 L -> A (IN REF. 2).  
 FT CONFLICT 150 150 P -> A (IN REF. 2).  
 FT CONFLICT 204 204 M -> V (IN REF. 2).  
 FT CONFLICT 409 409 R -> A (IN REF. 2).  
 SQ SEQUENCE 1388 AA; 60822AD925FA093D CRC64;  
 \*Query Match 55.8%; Score 193; DB 1; Length 1388;  
 \*Best Local Similarity 49.2%; Pred. No. 8.8e-15;  
 Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0;  
 QY 1 ADRAAVPIYNKDELIFPSWEALFSGSEGPLKPGARIFSPDGKDVLRHTVTPQKSWHGS 60  
 DB 1268 AERYSLPIVNLKGQVLFNNMDSIFSGHGQFNMHPIPIYSFGDRDITWDSWPKVIMHGS 1327  
 QY 61 DPN 63  
 DB 1328 SPH 1330  
 RESULT 4  
 CB21\_SINUL STANDARD; PRT; 266 AA.  
 ID CB21\_SINUL  
 AC P13851;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Chlorophyll A-B binding protein 1, chloroplast precursor (LHCII type I  
 CAB-1) (LHCP).  
 GN CAB1.  
 OS Sinapis alba (White mustard) (Brassica hirta).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Sinapis.  
 OC NCBI\_Taxid=3728;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cotyledon;

RX MEDLINE=92322955; PubMed=1623179;  
 RA Gaudy A., Batschauer A., von Arnim A., Koesse H.,  
 RT "Isolation and characterization of a gene encoding a chlorophyll a/b-  
 binding protein from mustard and the targeting of the encoded protein  
 to the thylakoid membrane of pea chloroplasts in vitro.",  
 RL Plant Mol. Biol. 19:277-287 (1992).  
 CC  
 CC -1- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT  
 RECEPTOR, IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS  
 WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN  
 EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF  
 GRANAL MEMBRANES AND PHOTOREGULATED BY REVERSIBLE PHOSPHORYLATION  
 OF ITS THREONINE RESIDUES. BOTH ARE BELIEVED TO MEDIATE THE  
 DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.  
 CC  
 CC -1- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS (A & B) AND  
 -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X15894; CAA33903.1; -  
 CC EMBL: X16436; CAA34459.1; -  
 CC PIR: S22511; S22511.  
 DR InterPro: IPR001344; Chloro\_Abbind.  
 DR Pfam: PF00504; chloroa\_b-bind\_1.  
 DR ProDom: PD000275; Chloro\_Abbind; 1.  
 KW Chlorophyll; Photosynthesis; Photosystem I; Photosystem II;  
 KW Thylakoid; Membrane; Chloroplast; Transist peptide; Multigene family;  
 KW Transmembrane; Phosphorylation.  
 FT TRANSIT 1 34 CHLOROPLAST (PROBABLE).  
 FT CHAIN 35 266 CHLOROXYLL A-B BINDING PROTEIN 1.  
 FT TRANSMEM 99 118 POTENTIAL.  
 FT TRANSMEM 151 171 POTENTIAL.  
 FT TRANSMEM 220 236 POTENTIAL.  
 SQ SEQUENCE 266 AA; 28232 MW; 63BCD583DB46AB7C CRC64;  
 \*Query Match 18.4%; Score 63.5; DB 1; Length 266;  
 \*Best Local Similarity 39.0%; Pred. No. 2.9;  
 Matches 16; Conservative 4; Mismatches 18; Indels 3; Gaps 1;  
 QY 24 PFGSEGPLKPGARIFSPDGKDVLR---HPTVTPQKSWHGS 61  
 DB 13 PFGKAVKLSPGASVFGTGRVTKVTPGTPSGSPWIGSD 53  
 RESULT 5  
 ACS1\_KLUUA STANDARD; PRT; 683 AA.  
 ID ACS1\_KLUUA  
 AC O60011;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Acetyl-coenzyme A synthetase 1 (EC 6.2.1.1) (Acetate--CoA ligase 1)  
 DE (Acyl-activating enzyme 1).  
 GN ACS1.  
 OS Kluyveromyces fragilis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OC NCBI\_Taxid=28985;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140;  
 RC Zeeman A.M., Henkel C., Steensma H.Y.;  
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY BE REQUIRED FOR ASSIMILATION OF ETHANOL AND ACETATE.  
 CC -1- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +  
 CC acetyl-CoA.  
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME

FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, A061265; AAC16713.1; -  
 DR InterPro: IPR000873; AMP-bind.  
 DR Pfam: PF00501; AMP-binding; 1.  
 DR PROSITE: PS00455; AMP BINDING; 1.  
 KW Ligase; Multigene family.  
 SQ SEQUENCE 683 AA; 76029 MW; 7AC88764A97966DB CRC64;

Query Match 17.6%; Score 61; DB 1; Length 683;  
 Best Local Similarity 27.1%; Pred. No. 16;  
 Matches 16; Conservative 8; Mismatches 13; Indels 22; Gaps 3;

-----

OY 25 SGSEGPLKPGARIFSGDKGV-LRHPT-----WPG--KSWHGSQ 61  
 Db 460 AGGATPMKFGAALPFPFGIDLAVIDPTGTIGTGEHAGVLAIKRPMSPFARTIMKND 518

RESULT 6

ID PRTS\_SERMA STANDARD; PRT; 1045 AA.

AC P09489;

DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Extracellular serine protease precursor (EC 3.4.21.-).  
 OS *Serratia marcescens*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Serratia.  
 OX NCBI\_TaxId=615;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-35 AND 407-408.  
 RC STRAIN=IFO 3046;  
 RA MEDLINE=86223815; PubMed3011754;  
 RA Yanagida N., Uozumi T., Beppu T.;  
 RT "Specific excretion of Serratia marcescens protease through the outer  
 RT membrane of *Escherichia coli*.";  
 RL J. Bacteriol. 166:937-944 (1986).  
 RP PARTIAL SEQUENCE, AND PROCESSING.  
 RX MEDLINE=92348352; PubMed1639760;  
 RA Shikata S., Shimada K., Kataoka H., Hornouchi S., Beppu T.;  
 RT "Detection of large COOH-terminal domains processed from the  
 RT precursor of Serratia marcescens serine protease in the outer  
 RT membrane of *Escherichia coli*.";  
 RL J. Biochem. 111:637-632 (1992).  
 CC -1- SUBCELLULAR LOCATION: secreted.  
 CC -1- SIMILARITY: Belongs to peptidase family S8.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, M13469; AA26572.1; -  
 DR PIR, A29840; A29840.  
 DR HSSP, Q99405; IMPT.  
 DR MEROPS, S08\_094; -  
 DR InterPro: IPR006315; Autotransport.  
 DR InterPro: IPR005546; Autotransporter.  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR Pfam: PF03797; Autotransporter; 1.

DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRFAMs; TIGR01414; autotrans\_bar1; 1.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; FALSE\_NEG.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KW Hydrolase; Serine protease; Zymogen; Signal.  
 FT SIGNAL 1  
 FT CHAIN 28 645 EXTRACELLULAR SERINE PROTEASE.  
 FT PROPEP 646 1045  
 FT ACT SITE 76 76 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT SITE 341 341 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 SQ SEQUENCE 1045 AA; 112345 MW; 4924E450E4FF179C CRC64;

Query Match 17.6%; Score 61; DB 1; Length 1045;  
 Best Local Similarity 34.9%; Pred. No. 25;  
 Matches 15; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

OY 14 ELTFPSWEALFSGSGPLKPGARIFSGDKVLRHPTVPQSV 56  
 Db 449 ECSFDSWNSDISGHGGLTKTGAGTLALGNNTYRQDTWVKQV 491

RESULT 7

ID TKT1\_RHIME STANDARD; PRT; 694 AA.

AC P58333;

DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Transketolase (EC 2.2.1.1) (TK).  
 GN CBRT OR RB0193 OR SMD2020.  
 OS *Rhizobium meliloti* (Sinorhizobium meliloti).  
 OC Plasmid pSymB (megaplasmid 2).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
 OX NCBI\_TaxId=382;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RC MEDLINE=21396508; PubMed11481431;  
 RA Finan T.M., Weidner S., Wong K., Buhrmaster J., Chain P.,  
 RA Vorvorcelter F.J., Hernandez-Lucas I., Becker A., Gouzy J.,  
 RA Golding B., Pechler A.;  
 RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the  
 RT N2-fixing endosymbiont *Sinorhizobium meliloti*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).  
 CC -1- CATALYTIC ACTIVITY: Sedoheptulose 7-phosphate + D-glyceraldehyde  
 CC 3-phosphate = D-ribose 5-phosphate + D-xylulose 5-phosphate.  
 CC -1- COFACTOR: THIAMINE PYRROPHOSPHATE (By similarity).  
 CC -1- PATHWAY: Calvin cycle.  
 CC -1- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, AL603642; CAC68593.1; -  
 DR PIR, A95866; A95866.  
 DR InterPro: IPR005478; BactTransketolase.  
 DR InterPro: IPR005476; Transketolase\_C.  
 DR InterPro: IPR005475; Transketolase\_CR.  
 DR InterPro: IPR005474; Transketolase\_N.  
 DR Pfam: PF03779; transket\_pyr; 1.  
 DR Pfam: PF00456; transketolase; 1.  
 DR Pfam: PF02780; transketolase\_C; 1.  
 DR TIGRFAMs; TIGR00232; tk1ase\_bact; 1.  
 DR PROSITE; PS00801; TRANSKETOLASE\_1; 1.



```

CC (POTENTIAL).
CC -1- DOMAIN: Lacks the zinc protease domain of other flesh proteins.
CC -2- Also much longer in both the N- and C-terminal.
CC -3- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB001684; BAA57905.1; -.
CC EMBL: AB001684; BAA57906.1; ALT_INIT.
CC PIR: T07258; T07258.
CC InterPro: IPR003599; AAA_ATPase.
CC InterPro: IPR003959; AAA_ATPase_cent.
CC InterPro: IPR003960; AAA_gub.
CC InterPro: IPR006642; Peptidase_M41.
CC Pfam: PF00004; AAA; 1.
CC Pfam: PF01434; Peptidase_M41; 1.
CC SMART: SM00382; AAA; 1.
CC PROSITE: PS00674; AAA; 1.
CC Cell division; ATP-binding; Transmembrane; Chloroplast.
CC KW TRANSMEM 48 68 POTENTIAL.
CC FT TRANSMEM 896 916 POTENTIAL.
CC FT TRANSMEM 973 993 POTENTIAL.
CC FT TRANSMEM 1021 1041 POTENTIAL.
CC SEQUENCE 1720 AA; 197172 MW; 0P1EA926B799D5B8 CRC64;

Query Match
Best Local Similarity 26.7%; Pred. No. 86;
Matches 16; Conservative 14; Mismatches 21; Indels 9; Gaps 4;

QY 9 VNLKDELLP-PSWEALFSGSGEPLKP-----GARIFSGDKVLNHPW-PKSYWHS 60
DB 1480 IGFEDDIRYSSPMWQDVS-AEMEFKPKPKGSRILYLVYERTSNPEWVPDEHYHNS 1538

RESULT 10
ACSA_SCHPO STANDARD; PRT; 662 AA.
ID ACSA_SCHPO
AC P78773;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable acetyl-coenzyme A synthetase (EC 6.2.1.1) (Acetate--CoA
DE ligase) (Acyl-activating enzyme).
OS SPC1191.02C OR SPC417.14C.
GN Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
OK [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones W., Leach S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymourez B.,
RA Weljens I., Vanetseels E., Rieger M., Schaefer M., Mueller-Auer S.,

```

```

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelure V., Mottier S.,
RA Gilbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen O., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
RA The genome sequence of Schizosaccharomyces pombe.
RA Nature 415:871-880(2002).
RN [12]
RP SEQUENCE OF 97-662 FROM N.A.
RX STRAIN=PR745;
RX MEDLINE=98162722; PubMed=9501991;
RA Yoshida S., Kato K., Nakai K., Okayama H., Nojima H.,
RT Identification of open reading frames in Schizosaccharomyces pombe
RT cDNA.
RL DNA Res. 4:363-369(1997).
CC -1- CATALYTIC ACTIVITY: ATP + acetate + CoA = AMP + diphosphate +
CC acetyl-CoA.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL049644; CAB41048.1; -.
CC EMBL: AL035076; CAA22660.1; -.
CC EMBL: D89121; BAA13783.1; -.
CC PIR: T41215; T41215.
CC GeneDB: Spombe: SPC1191.02C; -.
CC InterPro: IPR000873; AMP-bind.
CC Pfam: PF00501; AMP-binding; 1.
CC PROSITE: PS04455; AMP_BINDING; 1.
CC KW Ligase.
CC SEQUENCE 662 AA; 73035 MW; 02FD119ABB622583 CRC64;

Query Match
Best Local Similarity 28.8%; Score 58; DB 1; Length 662;
Matches 17; Conservative 6; Mismatches 12; Indels 24; Gaps 4;

QY 26 GSEGLPKGARIKFSF-----DGKDV-----LRHPTWPO-KSYWHS 61
DB 435 GPVTMKKGSATLPFGIDAVYIDPLTGKTIIGNDVEGLAIRSF-WPSARTVRGHD 492

RESULT 11
FXC1_HUMAN STANDARD; PRT; 553 AA.
ID FXC1_HUMAN
AC Q12948; OSBYM1; OSNUS5; OSUDDO; OSUP06;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Forkhead box protein C1 (Forkhead-related protein FKHL7) (Forkhead-
DE related transcription factor 3) (FKHRC-3).
GN FOXO1 OR FKHL7 OR FKRC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;
OK [1]
RP SEQUENCE FROM N.A., AND VARIANTS SER-112; MET-126 AND LEU-131.
RX MEDLINE=96282091; PubMed=9620769;
RA Nishimura D.Y., Swideraki R.E., Alward W.L.M., Seaby C.C.,
RA Patil S.R., Bennett S.R., Kanis A.B., Gaetfieri J.M., Stone E.M.,
RA Sheffield V.C.;

```

RT "The forkhead transcription factor gene FKHL7 is responsible for  
RT glaucoma phenotypes which map to 6p25.";  
RL Nat. Genet. 19:140-147(1998).  
RN [2]  
RP SEQUENCE FROM N.A., AND VARIANTS ARA THR-82 AND MET-87.  
RX MEDLINE=99011252; PubMed=9792859;  
RA Mearns A.J., Jordan T., Mirzayans F., Dubois S., Kume T., Parlee M.,  
RA Ritch R., Koop B., Kuo W.-L., Collins C., Marshall J., Gould D.B.,  
RA Pearce W., Carlsson P., Enerbaeck S., Morissette J., Bhattacharya S.,  
RA Hogan B., Raymond V., Walter M.A.;  
RT "Mutations of the forkhead/winged-helix gene, FKHL7, in patients with  
RT Axenfeld-Rieger anomaly.";  
RL Am. J. Hum. Genet. 63:1316-1328(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Patel R.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 68-177 FROM N.A., AND TISSUE SPECIFICITY.  
RC TISSUE=lympholeukemia;  
RA MEDLINE=93271467; PubMed=8499623;  
RX Hromas R., Moore J., Johnston T., Socha C., Klemasz M.;  
RT "Drosophila forkhead homologues are expressed in a lineage-restricted  
RT manner in human hematopoietic cells.";  
RL Blood 81:2854-2859(1993).  
RN [5]  
RP SEQUENCE OF 73-178 FROM N.A.  
RX MEDLINE=95045392; PubMed=7957066;  
RA Pierron S., Hellyer M., Samuelson L., Enerbaeck S., Carlsson P.;  
RT "Cloning and characterization of seven human forkhead proteins:  
RT binding site specificity and DNA bending.";  
RL EMO J. 13:5002-5012(1994).  
CC -1- FUNCTION: BINDING OF PREAC-3 AND PREAC-4 TO THEIR COGNATE SITES  
CC RESULTS IN BENDING OF THE DNA AT AN ANGLE OF 80-90 DEGREES.  
CC -1- SUBUNIT: Monomer.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: Expressed in all tissues and cell lines  
CC examined.  
CC -1- DISEASE: DEFECTS IN FOXK1 ARE THE CAUSE OF A SPECTRUM OF GLAUCOMA  
CC PHENOTYPES SUCH AS AXENFELD-RIEGER ANOMALY (ARA), AXENFELD-RIEGER  
CC SYNDROME (ARS) AND IRIDOGONIOSGENESIS ANOMALY (IGDA). ARS IS AN  
CC AUTOSOMAL DOMINANT DISORDER PRESENTING WITH ARA-LIKE OCULAR  
CC FINDINGS IN ADDITION TO ABNORMALITIES OF THE TEETH, JAW AND  
CC UMBILICUS.  
CC -1- SIMILARITY: Contains 1 fork-head domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
CC -----  
CC EMBL, AF046693; AAC18081.1; -;  
CC EMBL, AF078096; AAC72915.1; -;  
CC EMBL, AL034344; CAB81658.1; -;  
CC EMBL, L12143; AAK13575.1; -;  
CC EMBL, U13221; AAA92038.1; -;  
CC PIR, S51626; S51626.  
CC HSSP, O63245; 2HFM.  
CC TRANSFAC, T02471; -;  
CC Genew, HGNC:3800; FOXK1.  
CC MIM, 601631; -;  
CC DR GO, GO:0005634; Cnucleus; NAS.  
CC DR GO, GO:0003700; P:transcription factor activity; TAS.  
CC DR GO, GO:0016563; P:transcriptional activator activity; TAS.  
CC DR GO, GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
CC DR GO, GO:0007601; P:vision; TAS.  
CC DR InterPro, IPR001766; TF\_Fork\_head.  
CC DR Pfam, PF00250; Fork\_head\_1.  
CC PRINTS, PR00053; FORKHEAD.

DR Prodom; P000425; TF\_Fork\_head; 1.  
DR SMART, SM00339; FH; 1.  
DR PROSITE, PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE, PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE, PS0039; FORK\_HEAD\_3; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation;  
KW Disease mutation.  
FT DNA BIND 77 168  
FT DOMAIN 28 33 FORK-HEAD.  
FT DOMAIN 159 173 POLY-ALA.  
FT DOMAIN 194 197 POLY-ARG.  
FT DOMAIN 262 272 POLY-PRO.  
FT DOMAIN 292 297 POLY-SER.  
FT DOMAIN 375 382 POLY-GLY.  
FT DOMAIN 438 445 POLY-SER.  
FT DOMAIN 447 456 POLY-GLY.  
FT DOMAIN 486 495 POLY-ALA.  
FT VARIANT 82 82  
FT VARIANT 87 87  
FT VARIANT 112 112  
FT VARIANT 126 126  
FT VARIANT 131 131  
FT CONFLICT 70 77  
FT CONFLICT 101 101  
FT CONFLICT 180 180  
FT CONFLICT 199 202  
FT CONFLICT 426 426  
SQ SEQUENCE 553 AA; 56788 MW; 596F94303BE59A CAC64;  
Query Match 16.6%; Score 57.5; DB 1; Length 553;  
Best Local Similarity 37.9%; Pred. No. 33;  
Matches 22; Conservative 2; Mismatches 29; Indels 5; Gaps 2;  
QY 4 AAVPVLNKLDELPPSWALPSSGSGCP-LKGGARIFSPDGKDYLRHPWPQKSVHGS 60  
DB 232 AAVPRLSPDS-----SSSLSSGSSPPGSLPRLSLDGADSDAPPAPAPPPHHS 305  
RESULT 12  
KCEL\_MOUSE  
ID KCEL\_MOUSE STANDARD; PRT; 693 AA.  
AC P23298;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Protein kinase C, eta type (BC 2.7.1.-) (PKC-eta) (PKC-L).  
GN PRKCH OR PKCH.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Epidermis; PubMed=2266135;  
RX MEDLINE=91093089;  
RA Osada S.I., Mizuno K., Saido T.C., Akita Y., Suzuki K., Kuroki T.,  
RA Ohno S.;  
RT "A phorbol ester receptor/protein kinase, nPKC eta, a new member of  
RT the protein kinase C family predominantly expressed in lung and  
RT skin.";  
RL J. Biol. Chem. 265:22434-22440(1990).  
CC -1- FUNCTION: THIS IS CALCIUM-INDEPENDENT, PHOSPHOLIPID-DEPENDENT,  
CC SERINE- AND THREONINE-SPECIFIC ENZYME.  
CC -1- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN  
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS  
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.  
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LUNG AND SKIN.  
CC -1- SIMILARITY: Contains 2 zinc-dependent phorbol-ester and DAG







Best Local Similarity 30.6%; Pred. No. 62;  
Matches 15; Conservative 7; Mismatches 26; Indels 1; Gaps 1;

Qy 2 DRAAIVPVLKDELFPMSWEALFSGSGEPLKPGARIFSPDGKVLPHPT 50  
Db 572 DVVALPISQVVRSTPKPASQALHTSQSPRATSVYEDHSAK-ILRHPS 619

RESULT 14  
ID TP67 MYCTU

AC OS0654; OS0731; STANDARD; PRT; 884 AA.

DT 01-NOV-1997 (Rel. 35, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DN Hypothetical protein RV2567.

GN RV2567 OR MT2643 OR MTCY227.34C OR MTCY9C4.01C.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI\_TaxID=1773;

RP [1]

RC STRAIN=H37RV;

RA MEDLINE=98295987; PubMed=9634230.

RA Cole S.T., Broesch R., Parthill U., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Björmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock R., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA Hornby T., Jagals K., Kiroch A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,

RA Rutter S., Taylor K., Whitehead S., Barrell B.G.;

RA "Deciphering the Biology of Mycobacterium tuberculosis from the

RT complete genome sequence.";

RL Nature 393:537-544(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA Bishai W.;

RA "Whole genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains.";

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: SOME, TO M.TUBERCULOSIS RV2411C AND SYNECHOCYSTIS PCC

CC 6803 SLI0335.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z77250; CAB01053.1;

CC EMBL; AE007098; AAK46956.1;

CC PIR; C70729; C70729.

CC TIGR; MT2643;

CC Tuberculin; RV2567;

CC Pfam; PF04168; DUF403; 1.

CC Pfam; PF04169; DUF404; 1.

CC Pfam; PF04174; DUF407; 1.

CC Hypothetical protein; Complete proteome.

CC CONFLICT 645 645 O -> R (IN REF. 2).

CC SEQUENCE 884 AA; 95448 MW; 95023A4D2EDB365 CRC64;

CC Query Match 16.3%; Score 56.5; DB 1; Length 884;

CC Best Local Similarity 30.2%; Pred. No. 72;

CC Matches 16; Conservative 11; Mismatches 13; Indels 13; Gaps 2;

Qy 23 LFSGSGPLKPGARIFSPDG-----KDVLRHPT-----PKSWHSGSDP 62  
Db 620 LMSLTVDPDRPSGLVOSVSGALAAQAVRDLSNDTMMVLANVERAVERHKSDDP 672

RESULT 15  
ID SYM METUA

AC OS8659; STANDARD; PRT; 651 AA.

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Methionyl-L-cysteine synthetase (EC 6.1.1.10) (Methionine--cysteine ligase)

DE (MethS)

GN METG OR MJ1263.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;

OC Methanocaldococcaceae; Methanocaldococcus.

OX NCBI\_TaxID=2190;

RP [1]

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RA MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake O., FitzGerald L.M., Clayton R.A., Gocayne J.D.,

RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,

RA Scott J.L., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,

RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Huzar M.A., Kaine B.P., Borodovsky M.,

RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;

RA "Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii.";

RL Science 273:1058-1073(1996).

CC -1- FUNCTION: Is required not only for elongation of protein synthesis

CC but also for the initiation of all mRNA translation through

CC initiator tRNA (Met) aminoacylation (By similarity).

CC -1- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA(Met) = AMP +

CC L-methionyl-L-cysteine (Met).

CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

CC -1- SUBUNIT: Homodimer (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

CC MetG subfamily 1.

CC -1- SIMILARITY: Contains 1 tRNA-binding domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U67567; AAB99269.1;

CC PIR; F64457; F64457.

CC HSSP; P00959; IMEA.

CC TIGR; MJ1263;

CC HAMAP; MF\_00098; fused. 1.

CC InterPro; IPR004495; MetG\_Cterm.

CC InterPro; IPR002300; tRNA-synt\_1a.

CC InterPro; IPR001412; tRNA-synt\_1.

CC InterPro; IPR002304; tRNA-synt\_met.

CC InterPro; IPR002547; tRNA-synt.

CC Pfam; PF00133; tRNA-synt\_1; 1.

CC Pfam; PF01588; tRNA\_bind; 1.

CC PIRSF; PIRSF001528; MetG\_dimerising; 1.

CC TIGRFAMs; TIGR00398; metG; 1.

CC TIGRFAMs; TIGR00399; metG\_Cterm; 1.

CC PROSITE; PS00178; AA tRNA\_LIGASE\_I; FALSE\_NEG.

CC PROSITE; PS50886; TRND; 1.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 14:09:48 ; Search time 21 Seconds

(without alignments)  
288.506 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346  
Sequence: 1 ADRAAVPIVNLKDELLPSPW.....DVLNHTWPQKSVHSGDPN 63

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: PIR 761\*

1: PIR1\*  
2: PIR2\*  
3: PIR3\*  
4: PIR4\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	346	100.0	684	2 AS5019	collagen alpha 1(X
2	294	85.0	1774	2 B56101	collagen alpha 1(X
3	291	84.1	1315	2 AS6101	collagen alpha 1(X
4	193	55.8	1388	2 AS3317	collagen alpha 1(X
5	140	40.5	650	2 T22002	hypothetical prote
6	67.5	19.5	614	2 S27962	modulator recognit
7	64	18.5	644	2 A72519	probable 2-oxoacid
8	63.5	18.4	266	2 S22511	chlorophyll a/b-bi
9	63	18.2	336	2 T17408	RNA (adenine-N6-)
10	62.5	18.1	467	2 T26705	hypothetical prote
11	62.5	18.1	620	2 T68797	conserved hypotet
12	62	17.9	419	2 T25565	hypothetical prote
13	61	17.6	1045	2 A28840	serine proteinase
14	60	17.3	694	2 A95866	probable transkeo
15	60	17.3	984	2 T44496	cellulose 1,4-beta
16	59.5	17.2	268	2 H75614	hypothetical prote
17	59.5	17.2	362	2 S24434	class I histocompa
18	59	17.1	303	2 T41056	conserved hypotet
19	58.5	16.9	305	2 S07115	class I histocompa
20	58.5	16.9	354	2 I80170	class I histocompa
21	58.5	16.9	481	2 T23131	hypothetical prote
22	58.5	16.9	1184	2 D63387	probable protein p
23	58.5	16.9	1720	2 T07258	cell division prot
24	58	16.8	461	2 AS4024	protein kinase (EC
25	58	16.8	662	2 T41215	probable acetate-C
26	58	16.8	777	2 AD0982	biclin sulfoxide r
27	58	16.8	2055	2 T30259	multiple PDZ domai
28	57.5	16.6	506	2 H64618	sigma54 interacti
29	57.5	16.6	732	2 AF0439	probable Rhs acces

30	57.5	16.6	800	2 AG0438	probable Rhs acces
31	57	16.5	633	2 A23174	hypothetical prote
32	57	16.5	385	1 A23690	protein kinase (EC
33	57	16.5	1116	2 B88612	protein Y6A2A.2 (
34	57	16.5	1238	1 JC5573	copper-transportin
35	57	16.5	2054	2 T46612	multi PDZ domain p
36	56.5	16.3	211	2 T31724	hypothetical prote
37	56.5	16.3	217	2 T27331	hypothetical prote
38	56.5	16.3	256	2 G35866	probable pyrroloqu
39	56.5	16.3	770	2 S00643	anthranilate synth
40	56.5	16.3	884	2 C70729	hypothetical prote
41	56	16.2	651	2 F54457	methionine-tRNA li
42	56	16.2	753	1 D72660	probable aldehyde
43	56	16.2	838	2 T45699	hypothetical prote
44	55.5	16.0	205	2 F87660	hydrolyase, haloaci
45	55.5	16.0	299	2 D75138	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

AS5019  
collagen alpha 1(XVIII) chain - human (fragment)

N/Contains: endostatin

C/Species: Homo sapiens (man)

C/Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 31-Mar-2000

C/Accession: AS5019

R/Oh: S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen, Genomics 19, 494-499, 1994

A/Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localiza

A/Reference number: AS5019; MUID:9424537; PMID:8188291

A/Accession: AS5019

A/Molecule type: mRNA

A/Residues: 1-684 <OH>

A/Cross-references: GB:L22548; NID:G348908; PIDN:AA51864.1; PID:G562794

A/Note: The cited accession number, L22548, is not in Genbank release 103

A/Note: In the authors' translation, 482-61y is not shown, residues 483-490 are shifted

C/Comment: Proline and lysines at the third position of the tripeptide repeating unit (

lated and subsequently O-glycosylated.

C/Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivasc

C/Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un

C/Genetics:

A/Gene: GDB:COL18A1

A/Cross-references: GDB:138752; OMIM:120328

A/Map position: 21q22.3-21q22.3

C/Superfamily: unassigned collagens

C/Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc

F.1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>

F.1-59/Domain: collagenous (fragment) #status predicted <CO4>

F.74-115/Domain: collagenous #status predicted <CO5>

F.129-201/Domain: collagenous #status predicted <CO6>

F.212-244/Domain: collagenous #status predicted <CO7>

F.257-278/Domain: collagenous #status predicted <CO8>

F.262-264/Region: cell attachment (R-G-D) motif

F.286-340/Domain: collagenous #status predicted <CO9>

F.354-371/Domain: collagenous #status predicted <CO10>

F.502-684/Product: endostatin #status predicted <EST>

F.509-684/Region: multiplexin collagen carboxyl-terminal similarity

##### Query Match

Best Local Similarity 100.0%; Pred. No. 9.9e-34;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ADRAAVPIVNLKDELLPSPWALPSGEGPLKPGARIFSPGKQVLRHPPTWPQKSVHSGS	60
DB	565	ADRAAVPIVNLKDELLPSPWALPSGEGPLKPGARIFSPGKQVLRHPPTWPQKSVHSGS	624
QY	61	DPN 63	
DB	625	DPN 627	

## RESULT 2

B56101

collagen alpha 1(XVII) chain precursor, long splice form - mouse  
 Molecule type: collagen alpha 1(XVII) chain precursor, medium splice form; endostatin  
 C/Species: Mus musculus (house mouse)

C/Date: 03-Oct-1995 #sequence revision 08-May-1998 #text change 31-Mar-2000

C/Accession: B56101, C56101, S72450, S65595, P06751, A54072, A58816

R/Rehm, M.; Pihlajaniemi, T.

J. Biol. Chem. 270, 4705-4711, 1995

A/Title: Identification of three N-terminal ends of type XVII collagen chains and their  
 relation to homologous rat and Drosophila frizzled proteins.

A/Reference number: A56101; MUID:95181468; PMID:7876242

A/Accession: B56101

A/Molecule type: mRNA

A/Residues: 1-562 <REH1>

A/Cross-references: GB:U11637, NID:9618429, PIDN:AAC52179.1, PID:9618430

A/Experimental source: splice form clone PB17.24

A/Accession: C56101

A/Molecule type: mRNA

A/Residues: 1-239, 487-562 <REH2>

A/Cross-references: GB:U11637, NID:9618429

A/Experimental source: splice form clones PB8.1, PB19, PB15.2

R/Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.

submitted to the EMBL Data Library, August 1993

A/Reference number: S72450

A/Accession: S72450

A/Molecule type: mRNA

A/Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16

A/Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298

R/Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994

A/Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa

A/Reference number: A58370; MUID:94240111; PMID:8183893

A/Accession: S65595

A/Molecule type: mRNA

A/Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>

A/Cross-references: EMBL:L22545

R/Abde, N.; Muragaki, Y.; Yoshida, H.; Inoue, H.; Nishimura, Y.

Biochem. Biophys. Res. Commun. 196, 576-582, 1993

A/Title: Identification of a novel collagen chain represented by extensive interruptions

A/Reference number: P06751; MUID:94059075; PMID:8240330

A/Accession: P06751

A/Molecule type: mRNA

A/Residues: 635-1774 <ABE>

R/Rehm, M.; Hincik, E.; Pihlajaniemi, T.

J. Biol. Chem. 269, 13929-13935, 1994

A/Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str

collagen chain.

A/Reference number: A54072; MUID:94245707; PMID:8186673

A/Accession: A54072

A/Molecule type: DNA; mRNA

A/Residues: 1293-1403, 'R', 1405-1774 <REH3>

A/Cross-references: GB:U03714; NID:9487733; PIDN:AA20657.1; PID:9487734

R/O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukui, N.; Vasios, G.; Lane, W.S.; Flynn, E.; B

Cell 88, 277-285, 1997

A/Title: Endostatin, an endogenous inhibitor of angiogenesis and tumor growth.

A/Reference number: A58816; MUID:97160848; PMID:9008168

A/Accession: A58816

A/Molecule type: protein

A/Residues: 1591-1610 <ORE>

A/Experimental source: hemangioendothelial cells

A/Note: Inhibits endothelial cell proliferation

C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit

lated and subsequently O-glycosylated.

C/Comment: The different splice forms of collagen alpha 1(XVII) may be involved in per

C/Comment: Endostatin is released from collagen alpha 1(XVII) chain by the action of ur

ay be useful in treating solid tumors.

C/Genetics:

A/Genes: MGI:Coll18a1

A/Cross-references: MGI:71175

A/Map position: 10:91.0

A/Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599/

A/Note: the list of introns is incomplete

C/Superfamily: unassigned collagens

C/Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteogly

F1-1774/Product: collagen alpha 1(XVII) chain precursor, long splice form #status pred

F1-239, 487-1774/Product: collagen alpha 1(XVII) chain precursor, medium splice form #

F1-24/Dominant: signal sequence #status predicted <SIG>

F1-486/Region: frizzled similarity

F1-786-812/Dominant: collagenous #status predicted <CO01>

F1-823-896/Dominant: collagenous #status predicted <CO02>

F1-921-1042/Dominant: collagenous #status predicted <CO03>

F1-1066-1148/Dominant: collagenous #status predicted <CO04>

F1-1163-1204/Dominant: collagenous #status predicted <CO05>

F1-1218-1290/Dominant: collagenous #status predicted <CO06>

F1-1301-1333/Dominant: collagenous #status predicted <CO07>

F1-1346-1369/Dominant: collagenous #status predicted <CO08>

F1-1351-1353/Region: cell attachment (R-G-D) motif

F1-1377-1428/Dominant: collagenous #status predicted <CO09>

F1-1442-1459/Dominant: collagenous #status predicted <CO10>

F1-1591-1774/Product: endostatin #status predicted <EST>

F1-1598-1774/Region: multiplexin collagen carboxyl-terminal similarity

F1-354, 361, 947/Binding site: carbohydrate (Asn) (covalent) #status predicted

F1-699, 704, 1716/Binding site: carbohydrate (Ser) (covalent) #status predicted

F1-910, 913, 1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

## Query Match

Best Local Similarity 85.0%; Score 294; DB 2; Length 1774;

Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSGSECPKPGARIFSPOCKDYLRRPTMPQKSVHMG 60

DB 1654 ADRGVSPIVNLKDEVLSPSMDLFSGSGOGLPGARIFSPOCKDYLRRPTMPQKSVHMG 1713

QY 61 DPN 63

DB 1714 DPS 1716

## RESULT 3

collagen alpha 1(XVIII) chain precursor, short splice form - mouse

N/Contents: endostatin

C/Species: Mus musculus (house mouse)

C/Date: 03-Oct-1995 #sequence revision 08-May-1998 #text change 31-Mar-2000

C/Accession: A56101; A58371; S72450; S65595

R/Rehm, M.; Pihlajaniemi, T.

J. Biol. Chem. 270, 4705-4711, 1995

A/Title: Identification of three N-terminal ends of type XVIII collagen chains and their

relation to homologous rat and Drosophila frizzled proteins.

A/Reference number: A56101; MUID:95181468; PMID:7876242

A/Accession: A56101

A/Molecule type: mRNA

A/Residues: 1-103 <REH1>

A/Cross-references: GB:U11636; NID:9618427; PIDN:AAC52178.1; PID:9618428

R/Rehm, M.; Pihlajaniemi, T.

Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994

A/Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous

A/Reference number: A58371; MUID:94240112; PMID:8183894

A/Accession: A58371

A/Molecule type: mRNA

A/Residues: 1-928 <REH2>

A/Cross-references: GB:L16898; NID:9404754; PIDN:AA37434.1; PID:9553894

R/Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.

submitted to the EMBL Data Library, August 1993

A/Reference number: S72450

A/Accession: S72450

A/Molecule type: mRNA

A/Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHM>

A/Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298

R/Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994

A/Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa

A/Reference number: A58370; MUID:94240111; PMID:8183893

A/Accession: S65595

A:Molecule type: mRNA  
 A:Residues: 28-1315 <OHS>  
 A:Cross-references: EMBL:122545  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit labeled and subsequently O-glycosylated.  
 C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in pericyte and endothelial cell release from collagen alpha 1(XVIII) chain by the action of an enzyme useful in treating solid tumors.  
 C:Genetics:  
 A:Gene: MGI:COL18A1  
 A:Cross-references: MGI:71175  
 A:Map position: 10:41.0  
 C:Superfamily: unassigned collagens  
 C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglycan  
 F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:26-1315/Region: chondroitin amino-terminal similarity  
 F:1327-353/Domain: collagen alpha 1(XVIII) chain, short splice form #status predicted <M>  
 F:1364-437/Domain: collagenous #status predicted <CO1>  
 F:1462-583/Domain: collagenous #status predicted <CO2>  
 F:1607-689/Domain: collagenous #status predicted <CO3>  
 F:1704-745/Domain: collagenous #status predicted <CO5>  
 F:1759-831/Domain: collagenous #status predicted <CO6>  
 F:1842-874/Domain: collagenous #status predicted <CO7>  
 F:1887-910/Domain: collagenous #status predicted <CO8>  
 F:1892-894/Region: cell attachment (R-G-D) motif  
 F:1918-969/Domain: collagenous #status predicted <CO9>  
 F:1983-1000/Domain: collagenous #status predicted <CO10>  
 F:1133-1315/Product: endostatin #status predicted <EST>  
 F:1126-488/Binding site: multiplexin collagen carboxyl-terminal similarity  
 F:1172-228/Disulfide bonds: #status predicted  
 F:1240-245/Binding site: carbohydrate (Ser) (covalent) #status predicted  
 F:1451-454/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match  
 Best Local Similarity 84.1%; Score 291; DB 2; Length 1315;  
 Matches 50; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAVPYVNLKDELFPSEMEALFSGSEGPLKRGARIFSPDGKDLRHPTWPKSVWHS 60  
 DB 1195 ADRGSPYVNLKDELFPSEMEALFSGSEGPLKRGARIFSPDGKDLRHPTWPKSVWHS 1254

QY 61 DPN 63  
 DB 1255 DPS 1257

RESULT 4  
 A53317  
 collagen alpha 1(XV) chain precursor - human  
 M:Alternate names: procollagen alpha 1(XV) chain  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 31-Mar-2000  
 C:Accession: A53317; A53146; S28778  
 R:Kivirikko, S.; Heinemann, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T.  
 J. Biol. Chem. 269, 4773-4779, 1994  
 A:Title: Primary structure of the alpha1 chain of human type XV collagen and exon-intron  
 A:Reference number: A53317; MUID:94148920; PMID:8106446  
 A:Accession: A53317  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1388 <KIV>  
 A:Cross-references: GB:125280  
 A:Note: nucleotide sequence and conceptual translation not complete  
 R:Muragaki, Y.; Abe, N.; Nimmiya, Y.; Olsen, B.R.; Ooshima, A.  
 J. Biol. Chem. 269, 4042-4046, 1994  
 A:Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triple  
 A:Reference number: A53146; MUID:94140817; PMID:8307960  
 A:Accession: A53146  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-9, 'S', 11-48, 'V', 50-94, 'A', 96-149, 'A', 151-203, 'V', 205-408, 'A', 410-569 <MUR>

A:Cross-references: GB:D21230; MUID:9415605; PIDN:BAA04762.1; PID:01005294; PID:9460703  
 R:Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992  
 A:Title: Identification of a previously unknown human collagen chain, alpha1(XV), charac  
 A:Reference number: S28778; MUID:9306196; PMID:1279671  
 A:Accession: S28778  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 544-640, 'P', 642-811, 'P', 813-1252 <MYE>  
 C:Genetics:  
 A:Gene: GDB:COL15A1  
 A:Cross-references: GDB:132578; OMIM:120325  
 A:Map position: 9q21-qg22  
 C:Superfamily: unassigned collagens  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:123-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>  
 F:1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match  
 Best Local Similarity 55.8%; Score 193; DB 2; Length 1388;  
 Matches 31; Conservative 16; Mismatches 16; Indels 0; Gaps 0;

QY 1 ADRAVPYVNLKDELFPSEMEALFSGSEGPLKRGARIFSPDGKDLRHPTWPKSVWHS 60  
 DB 1268 AERVSLPYVNLKQVTLFNNWDSIFSGHGQFNMHPIVSPDGRDYMTPSPQKVIWHS 1327

QY 61 DPN 63  
 DB 1328 SPH 1330

RESULT 5  
 T22002  
 hypothetical protein F39H1.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T22002  
 R:White, S.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19500  
 A:Accession: T22002  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-650 <WIL>  
 A:Cross-references: EMBL:Z81079; PIDN:CAB03084.1; GSPDB:GN00019; CESP:F39H1.4  
 A:Experimental source: clone F39H1  
 C:Genetics:  
 A:Gene: CESP:F39H1.4  
 A:Map position: 1  
 A:Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match  
 Best Local Similarity 40.5%; Score 140; DB 2; Length 650;  
 Matches 25; Conservative 9; Mismatches 17; Indels 2; Gaps 1;

QY 8 IVNLDELFPSEMEALFSGSEGPLKRGARIFSPDGKDLRHPTWPKSVWHS 60  
 DB 525 VVNVAGHLLFSPKRSFVNGAG--MPPAKLTSFSDHVDLNSRNPDKKVMHS 575

RESULT 6  
 S27962  
 modulator recognition factor 1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 05-Nov-1999  
 C:Accession: S27962  
 R:Okada, T.; Merrill, B.W.; Huang, E.; Okada, T.; Tanaka, Y.; Gertson, P.; Itakura, K.  
 submitted to the EMBL Data Library, March 1991  
 A:Reference number: S27962  
 A:Accession: S27962  
 A:Molecule type: mRNA  
 A:Residues: 1-614 <OKA>  
 A:Cross-references: EMBL:M62324; MUID:q188683; PIDN:AAA6325.1; PID:q188684

[illegible]

**RESULT 9**

T17408  
rRNA (adenine-N6-)-methyltransferase (EC 2.1.1.48) 1 - Streptomyces venezuelae C/Species: Streptomyces venezuelae  
C/Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 08-Sep-2000  
C/Accession: T17408  
R.Xue, Y./ Zhao, L./ Liu, H.W., Sherman, D.H.  
Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998  
A>Title: A gene cluster for macroide antibiotic biosynthesis in streptomyces venezuelae  
A/Reference number: Z18773; MUID:58445333; PMID:9770448  
A/Accession: T17408  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-336 <XUE>  
A/Cross-references: EMBL:AF079138; NID:g3808326; PID:g3808033; PIDN:AAC69328.1  
C/GeneticB:  
A/Gene: pIKR1  
C/Keywords: methyltransferase; S-adenosylmethionine

**Query Match** 18.2%; Score 63; DB 2; Length 336;  
**Best Local Similarity** 40.0%; Pred. No. 7.1;  
**Matches** 14; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLKDELFPSSWEALFSGSGGLKCA 35  
||| : | : ||| :  
Db 235 ADRHGIPDGGLPKDILTOWIALFOASQPSYPARGA 269

**RESULT 10**

T26705  
hypothetical protein Y38H6C.17 - Caenorhabditis elegans C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 12-May-2003  
C/Accession: T26705  
R.White, S.  
submitted to the EMBL Data Library, September 1998  
A/Reference number: Z20255  
A/Accession: T26705  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-467 <WIL>  
A/Cross-references: EMBL:AL031630; PIDN:CMA20995.1; GSDB:GN00023; CESP:Y38H6C.17  
A/Experimental source: clone Y38H6C  
C/Genetics:  
A/Gene: CESP:Y38H6C.17  
A/Map position: 5  
A/Linrome: 9/2; 48/2; 79/3; 109/2; 140/2; 168/3; 225/3; 322/2; 382/1; 429/2  
C/Superfamily: Arabidopsis amino acid transport protein 1

**Query Match** 18.1%; Score 62.5; DB 2; Length 467;  
**Best Local Similarity** 33.9%; Pred. No. 12;  
**Matches** 20; Conservative 12; Mismatches 12; Indels 15; Gaps 5;

OY 4 AAATIVVLKDELL---FPSMF-ALFGSGEG-PLKPGARIFSPDG-----KDVLRHP 49  
||| : | : ||| : | : ||| :  
Db 218 AAALIT--OELLSHTWPTMELPSTIGVGSLANGSLTISFEQQAMVPLENSLKH 274

**RESULT 11**

F69797  
conserved hypothetical protein yeeW - Bacillus subtilis C/Species: Bacillus subtilis  
C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C/Accession: F69797  
R.Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allioni, G.; Azevedo, V.; Berteiro  
C.; Bron, S.; Brulliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emmertson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using BW model

Run on: September 10, 2003, 14:10:33 ; Search time 26 Seconds

(without alignments)  
353.557 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346  
Sequence: 1 ADRAAVPIVNLKDELLFPSPW.....DVLRRHPTPKQKSVHMGSDPN 63

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 541936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTIS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10C\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	63	9	US-09-822-540A-1
2	346	100.0	178	15	US-10-131-241-60
3	346	100.0	178	15	US-10-042-347-5
4	346	100.0	179	15	US-10-131-241-57
5	346	100.0	180	15	US-10-131-241-56
6	346	100.0	181	15	US-10-131-241-55
7	346	100.0	182	10	US-09-998-831-14
8	346	100.0	182	15	US-10-131-241-54
9	346	100.0	182	15	US-10-042-347-3
10	346	100.0	183	9	US-09-873-676-2
11	346	100.0	183	12	US-10-282-418-4
12	346	100.0	183	14	US-10-080-797-1
13	346	100.0	183	15	US-10-131-241-52
14	346	100.0	684	11	US-09-951-403-5
15	346	100.0	1516	15	US-10-060-036-166

16	338	97.7	180	15	US-10-131-241-47	Sequence 47, Appl
17	302	87.3	184	12	US-10-292-418-35	Sequence 35, Appl
18	302	87.3	184	12	US-09-938-391-4	Sequence 4, Appl
19	302	87.3	184	15	US-10-131-241-49	Sequence 49, Appl
20	302	87.3	230	12	US-09-938-391-2	Sequence 2, Appl
21	294	85.0	191	10	US-09-998-831-13	Sequence 13, Appl
22	294	85.0	207	14	US-10-080-797-3	Sequence 3, Appl
23	291	84.1	184	12	US-10-292-418-18	Sequence 18, Appl
24	291	84.1	184	15	US-10-131-241-46	Sequence 46, Appl
25	277.5	80.2	185	14	US-10-036-869-36	Sequence 36, Appl
26	182	52.6	31	9	US-09-822-540A-2	Sequence 2, Appl
27	94	27.2	16	10	US-09-766-412-32	Sequence 32, Appl
28	67.5	19.5	614	15	US-10-146-473-45	Sequence 45, Appl
29	63	18.2	336	10	US-09-861-289-27	Sequence 27, Appl
30	63	18.2	336	10	US-09-860-846-27	Sequence 27, Appl
31	63	18.2	336	11	US-09-860-848-27	Sequence 27, Appl
32	63	18.2	336	11	US-09-836-821-27	Sequence 27, Appl
33	63	18.2	11877	10	US-09-861-289-6	Sequence 6, Appl
34	63	18.2	11877	10	US-09-860-846-6	Sequence 6, Appl
35	63	18.2	11877	11	US-09-836-821-6	Sequence 6, Appl
36	63	18.2	12199	11	US-09-988-384B-6	Sequence 6, Appl
37	62.5	18.1	620	15	US-10-124-880-6	Sequence 6, Appl
38	61.5	17.8	3170	15	US-10-128-714-8504	Sequence 8504, Ap
39	59	17.1	14	10	US-09-766-412-31	Sequence 31, Appl
40	57.5	16.6	506	12	US-09-862-227-116	Sequence 116, App
41	57.5	16.6	536	11	US-09-292-862-2	Sequence 2, Appl
42	57	16.5	636	15	US-10-156-761-13432	Sequence 13432, A
43	57	16.5	2037	10	US-09-951-401-3	Sequence 3, Appl
44	57	16.5	2037	10	US-09-922-101-3	Sequence 3, Appl
45	57	16.5	2037	10	US-09-951-402-3	Sequence 3, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-822-540A-1
; Sequence 1, Application US/09822540A
; Patent No. US20020058620A1
; GENERAL INFORMATION:
; APPLICANT: Phoenix Pharmaceuticals, Inc.
; TITLE OF INVENTION: Cell Growth Regulation System
; FILE REFERENCE: PhoenixPharFull1
; CURRENT APPLICATION NUMBER: US/09/822, 540A
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,561
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 63
; TYPE: PRT
; ORGANSIM: Homo sapiens
US-09-822-540A-1

Query Match      100.0%; Score 346; DB 9; Length 63;
Best Local Similarity 100.0%; Pred. No. 2.1e-35;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ADRAAVPIVNLKDELLFPSEWALFSGSEGLPKGARIFFPDGKDVLRHPTPKQKSVHMG 60
      |||
Db      1 ADRAAVPIVNLKDELLFPSEWALFSGSEGLPKGARIFFPDGKDVLRHPTPKQKSVHMG 60

Qy      61 DPN 63
      |||
Db      61 DPN 63

RESULT 2
US-10-131-241-60
; Sequence 60, Application US/10131241
; Publication No. US20030012792A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-60

Query Match          100.0%; Score 346; DB 15; Length 178;
Best Local Similarity 100.0%; Pred. No. 6.5e-35;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAPVIVNLKDELFPSWEALFSGSEGPLKPGARIFSPGKDVLRHPTWPKSVMHGS 60
DB 60 ADRAAPVIVNLKDELFPSWEALFSGSEGPLKPGARIFSPGKDVLRHPTWPKSVMHGS 119

QY 61 DPN 63
DB 120 DPN 122

RESULT 3
US-10-042-347-5
; Sequence 5, Application US/10042347
; Publication No. US20030114370A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael S.
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide Fr
; FILE REFERENCE: 05213-0880 (43170-249874)
; CURRENT APPLICATION NUMBER: US/10/042,347
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 09/315,689
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: US 60/106,343
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: US 09/154,302
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: US 08/740,168
; PRIOR FILING DATE: 1996-10-22
; PRIOR APPLICATION NUMBER: US 60/005,835
; PRIOR FILING DATE: 1995-10-23
; PRIOR APPLICATION NUMBER: US 60/023,070
; PRIOR FILING DATE: 1996-08-02
; PRIOR APPLICATION NUMBER: US 60/026,263
; PRIOR FILING DATE: 1996-09-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-347-5

Query Match          100.0%; Score 346; DB 15; Length 178;
Best Local Similarity 100.0%; Pred. No. 6.5e-35;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 ADRAAPVIVNLKDELFPSWEALFSGSEGPLKPGARIFSPGKDVLRHPTWPKSVMHGS 60
DB 60 ADRAAPVIVNLKDELFPSWEALFSGSEGPLKPGARIFSPGKDVLRHPTWPKSVMHGS 119

QY 61 DPN 63
DB 120 DPN 122

RESULT 4
US-10-131-241-57
; Sequence 57, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-57

Query Match          100.0%; Score 346; DB 15; Length 179;
Best Local Similarity 100.0%; Pred. No. 6.5e-35;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAPVIVNLKDELFPSWEALFSGSEGPLKPGARIFSPGKDVLRHPTWPKSVMHGS 60
DB 60 ADRAAPVIVNLKDELFPSWEALFSGSEGPLKPGARIFSPGKDVLRHPTWPKSVMHGS 119

QY 61 DPN 63
DB 120 DPN 122

RESULT 5
US-10-131-241-56
; Sequence 56, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-56
```

Query Match 100.0%; Score 346; DB 15; Length 180;  
Best Local Similarity 100.0%; Pred. No. 6,6e-35;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKGARIFSGDKDVLRRHTPWQKSWHGS 60  
DB 64 ADRAAVPIVNLKDELLFPGSEGLPKGARIFSGDKDVLRRHTPWQKSWHGS 123

QY 61 DPN 63  
124 DPN 126

RESULT 6  
US-10-131-241-55  
; Sequence 55; Application US/101331241  
; Publication No. US20030012792A1  
; GENERAL INFORMATION:  
; APPLICANT: Holaday, John W.  
; APPLICANT: Fortier, Anne H.  
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer  
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers  
; FILE REFERENCE: 05213-0344 43170-271565  
; CURRENT APPLICATION NUMBER: US/10/131,241  
; CURRENT FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: US 09/413,049  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: US 09/316,802  
; PRIOR FILING DATE: 1999-05-21  
; PRIOR APPLICATION NUMBER: US 60/086,586  
; PRIOR FILING DATE: 1998-05-22  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 55  
; LENGTH: 181  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-131-241-55

Query Match 100.0%; Score 346; DB 15; Length 181;  
Best Local Similarity 100.0%; Pred. No. 6,6e-35;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKGARIFSGDKDVLRRHTPWQKSWHGS 60  
DB 64 ADRAAVPIVNLKDELLFPGSEGLPKGARIFSGDKDVLRRHTPWQKSWHGS 123

QY 61 DPN 63  
124 DPN 126

RESULT 7  
US-09-998-831-14  
; Sequence 14; Application US/09998831  
; Patent No. US20020119153A1  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brecken  
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY  
; TITLE OF INVENTION: INHIBITING VEGF  
; FILE REFERENCE: 4001.002584  
; CURRENT APPLICATION NUMBER: US/09/998,831  
; CURRENT FILING DATE: 2001-11-30  
; PRIOR APPLICATION NUMBER: 09/561,108  
; PRIOR FILING DATE: 2000-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 182  
; TYPE: PRT  
; ORGANISM: Artificial Sequence

FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
; OTHER INFORMATION: PEPTIDE  
US-09-998-831-14

Query Match 100.0%; Score 346; DB 10; Length 182;  
Best Local Similarity 100.0%; Pred. No. 6,7e-35;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKGARIFSGDKDVLRRHTPWQKSWHGS 60  
DB 64 ADRAAVPIVNLKDELLFPGSEGLPKGARIFSGDKDVLRRHTPWQKSWHGS 123

QY 61 DPN 63  
124 DPN 126

RESULT 8  
US-10-131-241-54  
; Sequence 54; Application US/101331241  
; Publication No. US20030012792A1  
; GENERAL INFORMATION:  
; APPLICANT: Holaday, John W.  
; APPLICANT: Fortier, Anne H.  
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer  
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers  
; FILE REFERENCE: 05213-0344 43170-271565  
; CURRENT APPLICATION NUMBER: US/10/131,241  
; CURRENT FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: US 09/413,049  
; PRIOR FILING DATE: 1999-10-06  
; PRIOR APPLICATION NUMBER: US 09/316,802  
; PRIOR FILING DATE: 1999-05-21  
; PRIOR APPLICATION NUMBER: US 60/086,586  
; PRIOR FILING DATE: 1998-05-22  
; NUMBER OF SEQ ID NOS: 65  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 54  
; LENGTH: 182  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-131-241-54

Query Match 100.0%; Score 346; DB 15; Length 182;  
Best Local Similarity 100.0%; Pred. No. 6,7e-35;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKGARIFSGDKDVLRRHTPWQKSWHGS 60  
DB 64 ADRAAVPIVNLKDELLFPGSEGLPKGARIFSGDKDVLRRHTPWQKSWHGS 123

QY 61 DPN 63  
124 DPN 126

RESULT 9  
US-10-042-347-3  
; Sequence 3; Application US/10042347  
; Publication No. US20030114370A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Reilly, Michael S.  
; APPLICANT: Folkman, M. Judah  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide F  
; FILE REFERENCE: 05213-0880 (43170-249874)  
; CURRENT APPLICATION NUMBER: US/10/042,347  
; CURRENT FILING DATE: 2002-01-11  
; PRIOR APPLICATION NUMBER: US 09/115,689  
; PRIOR FILING DATE: 1998-05-20  
; PRIOR APPLICATION NUMBER: US 60/106,343  
; PRIOR FILING DATE: 1998-10-30

;; PRIOR APPLICATION NUMBER: US 09/154,302  
;; PRIOR FILING DATE: 1998-09-16  
;; PRIOR APPLICATION NUMBER: US 08/740,168  
;; PRIOR FILING DATE: 1996-10-22  
;; PRIOR APPLICATION NUMBER: US 60/005,835  
;; PRIOR FILING DATE: 1995-10-23  
;; PRIOR APPLICATION NUMBER: US 60/023,070  
;; PRIOR FILING DATE: 1996-08-02  
;; PRIOR APPLICATION NUMBER: US 60/026,263  
;; PRIOR FILING DATE: 1996-09-17  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 3  
;; LENGTH: 182  
;; TYPE: PRF  
;; ORGANISM: Homo sapiens  
US-10-042-347-3

Query Match 100.0%; Score 346; DB 15; Length 182;  
Best Local Similarity 100.0%; Pred. No. 6,7e-35;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSGSEGLPKPGARIFSPDGKDYLRHPTWPKSVHGS 60  
DB 64 ADRAAVPIVNLKDELLFSGSEGLPKPGARIFSPDGKDYLRHPTWPKSVHGS 123  
QY 61 DPN 63  
DB 124 DPN 126

## RESULT 10

US-09-873-676-2  
;; Sequence 2, Application US/09873676  
;; Patent No. US20020077289A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Macdonald, Nicholas J.  
;; APPLICANT: Sim, Kim L.  
;; TITLE OF INVENTION: Angiostatin and Endostatin Binding Proteins and Methods of Use  
;; FILE REFERENCE: 05213-0378 (43170-25333)  
;; CURRENT APPLICATION NUMBER: US/09/873,676  
;; PRIOR FILING DATE: 2001-06-04  
;; PRIOR APPLICATION NUMBER: US 60/209,065  
;; PRIOR FILING DATE: 2000-06-02  
;; PRIOR APPLICATION NUMBER: US 60/289,387  
;; PRIOR FILING DATE: 2001-05-08  
;; NUMBER OF SEQ ID NOS: 123  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 2  
;; LENGTH: 183  
;; TYPE: PRF  
;; ORGANISM: Homo sapiens  
US-09-873-676-2

Query Match 100.0%; Score 346; DB 9; Length 183;  
Best Local Similarity 100.0%; Pred. No. 6,7e-35;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSGSEGLPKPGARIFSPDGKDYLRHPTWPKSVHGS 60  
DB 64 ADRAAVPIVNLKDELLFSGSEGLPKPGARIFSPDGKDYLRHPTWPKSVHGS 123  
QY 61 DPN 63  
DB 124 DPN 126

RESULT 11  
US-10-292-418-4  
;; Sequence 4, Application US/10292418  
;; Publication No. US20030139365A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Lo, Kin-Ming

;; APPLICANT: Li, Yue  
;; APPLICANT: Gillies, Stephen D  
;; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as  
;; FILE REFERENCE: LEX-006C1  
;; CURRENT APPLICATION NUMBER: US/10/292,418  
;; PRIOR FILING DATE: 2002-11-12  
;; PRIOR APPLICATION NUMBER: 09/363,315  
;; PRIOR FILING DATE: 1999-08-25  
;; PRIOR APPLICATION NUMBER: US 60/097,883  
;; PRIOR FILING DATE: 1998-08-25  
;; NUMBER OF SEQ ID NOS: 54  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 4  
;; LENGTH: 183  
;; TYPE: PRF  
;; ORGANISM: Homo sapiens  
US-10-292-418-4

Query Match 100.0%; Score 346; DB 12; Length 183;  
Best Local Similarity 100.0%; Pred. No. 6,7e-35;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSGSEGLPKPGARIFSPDGKDYLRHPTWPKSVHGS 60  
DB 64 ADRAAVPIVNLKDELLFSGSEGLPKPGARIFSPDGKDYLRHPTWPKSVHGS 123  
QY 61 DPN 63  
DB 124 DPN 126

## RESULT 12

US-10-080-797-1  
;; Sequence 1, Application US/10080797  
;; Publication No. US20020183253A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Campochiaro, Peter A.  
;; APPLICANT: Brazzell, Romulus K.  
;; TITLE OF INVENTION: METHOD FOR TREATING OCULAR  
;; FILE REFERENCE: 4-31881A  
;; CURRENT APPLICATION NUMBER: US/10/080,797  
;; PRIOR FILING DATE: 2002-02-21  
;; NUMBER OF SEQ ID NOS: 21  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 1  
;; LENGTH: 183  
;; TYPE: PRF  
;; ORGANISM: Human  
US-10-080-797-1

Query Match 100.0%; Score 346; DB 14; Length 183;  
Best Local Similarity 100.0%; Pred. No. 6,7e-35;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSGSEGLPKPGARIFSPDGKDYLRHPTWPKSVHGS 60  
DB 64 ADRAAVPIVNLKDELLFSGSEGLPKPGARIFSPDGKDYLRHPTWPKSVHGS 123  
QY 61 DPN 63  
DB 124 DPN 126

RESULT 13  
US-10-131-241-52  
;; Sequence 52, Application US/10131241  
;; Publication No. US20030012792A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Holaday, John W.  
;; APPLICANT: Fortier, Anne H.

;; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Proliferation  
;; FILE REFERENCE: 05213-0344 43170-271565  
;; CURRENT APPLICATION NUMBER: US/10/131,241  
;; PRIOR FILING DATE: 2002-07-22  
;; PRIOR APPLICATION NUMBER: US 09/413,049  
;; PRIOR FILING DATE: 1999-10-06  
;; PRIOR APPLICATION NUMBER: US 09/316,802  
;; PRIOR FILING DATE: 1999-05-21  
;; PRIOR APPLICATION NUMBER: US 60/086,586  
;; PRIOR FILING DATE: 1998-05-22  
;; NUMBER OF SEQ ID NOS: 65  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO: 52  
;; LENGTH: 183  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-131-241-52

Query Match 100.0%; Score 346; DB 15; Length 183;  
Best Local Similarity 100.0%; Pred. No. 6.7e-35;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
DB 64 ADRAAVPIVNLKDELLFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 123

QY 61 DPN 63  
DB 124 DPN 126

RESULT 14  
US-09-961-403-5  
;; Sequence 5, Application US/09961403  
;; Publication No. US2003007589A1  
;; GENERAL INFORMATION:  
;; APPLICANT: HE-STUMP, HOLGER  
;; APPLICANT: HAENDLER, BERNARD  
;; APPLICANT: KRAETZSCHMAR, JOERN  
;; APPLICANT: KREFT, BERTHOLT  
;; APPLICANT: WINTERHAGER, ELKE  
;; APPLICANT: REGIDOR, PEDRO  
;; APPLICANT: SCOTTI, SIMONE  
;; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS  
;; FILE REFERENCE: SCH-1789  
;; CURRENT APPLICATION NUMBER: US/09/961,403  
;; CURRENT FILING DATE: 2001-09-25  
;; NUMBER OF SEQ ID NOS: 15  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO: 5  
;; LENGTH: 684  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-961-403-5

Query Match 100.0%; Score 346; DB 11; Length 684;  
Best Local Similarity 100.0%; Pred. No. 2.9e-34;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
DB 565 ADRAAVPIVNLKDELLFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 624

QY 61 DPN 63  
DB 625 DPN 627

RESULT 15  
US-10-060-036-166  
;; Sequence 166, Application US/10060036  
;; Publication No. US20030073144A1

;; GENERAL INFORMATION:  
;; APPLICANT: Benson, Darin R.  
;; APPLICANT: Kalos, Michael D.  
;; APPLICANT: Lodes, Michael J.  
;; APPLICANT: Persing, David H.  
;; APPLICANT: Hepler, William T.  
;; APPLICANT: Jiang, Yugu  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
;; FILE REFERENCE: 210121.566  
;; CURRENT APPLICATION NUMBER: US/10/060,036  
;; CURRENT FILING DATE: 2002-01-30  
;; NUMBER OF SEQ ID NOS: 4560  
;; SOFTWARE: FaastSeq for Windows Version 4.0  
;; SEQ ID NO: 166  
;; LENGTH: 1516  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-060-036-166

Query Match 100.0%; Score 346; DB 15; Length 1516;  
Best Local Similarity 100.0%; Pred. No. 7e-34;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
DB 1397 ADRAAVPIVNLKDELLFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVWHS 1456

QY 61 DPN 63  
DB 1457 DPN 1459

Search completed: September 10, 2003, 14:13:44  
Job time : 27 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 14:09:53; Search time 21 Seconds  
(without alignments)  
126.933 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346  
Sequence: 1 ADRAAVPIVNLKDELLFSPSW.....DYLHPTWPKSVHGDSPN 63

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents AA:  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6CTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	346	100.0	178	4	US-09-315-689-5
2	346	100.0	182	4	US-09-561-500-14
3	346	100.0	182	4	US-09-561-108-14
4	346	100.0	182	4	US-09-315-689-3
5	346	100.0	182	4	US-09-561-526-14
6	346	100.0	182	4	US-09-561-499-14
7	346	100.0	183	3	US-09-206-059-2
8	294	85.0	191	4	US-09-561-500-13
9	294	85.0	191	4	US-09-561-108-13
10	294	85.0	191	4	US-09-561-526-13
11	294	85.0	191	4	US-09-561-499-13
12	291	84.1	195	1	US-08-159-784-2
13	277.5	80.2	185	3	US-08-985-526-36
14	190	54.9	191	1	US-08-159-784-3
15	94	27.2	16	3	US-09-385-442-32
16	63	18.2	336	3	US-09-105-537-27
17	63	18.2	11877	3	US-09-105-537-6
18	62.5	18.1	620	4	US-09-311-626B-6
19	59.5	17.2	92	4	US-09-673-809-106
20	59	17.1	14	3	US-09-385-442-31
21	59	17.1	450	4	US-09-252-991A-16659
22	59	17.1	1044	4	US-09-252-991A-18853
23	57.5	16.6	506	2	US-08-849-480A-5
24	57.5	16.6	553	3	US-09-083-351-2
25	57.5	16.6	553	3	US-09-083-352-2
26	57	16.5	2037	3	US-09-306-998-3
27	55.5	16.0	191	4	US-09-376-728-2

28	55	15.9	444	1	US-08-483-140-28	Sequence 28, Appl
29	55	15.9	444	2	US-08-485-938A-32	Sequence 32, Appl
30	54.5	15.8	470	2	US-08-377-440A-1	Sequence 1, Appl
31	54.5	15.8	470	3	US-09-440-530-1	Sequence 1, Appl
32	54.5	15.8	578	1	US-08-653-740-3	Sequence 3, Appl
33	54.5	15.8	578	2	US-09-073-594-3	Sequence 3, Appl
34	54.5	15.8	578	3	US-09-275-925-3	Sequence 3, Appl
35	54.5	15.8	636	1	US-08-653-740-5	Sequence 3, Appl
36	54.5	15.8	636	2	US-09-073-594-5	Sequence 5, Appl
37	54.5	15.8	636	3	US-09-275-925-5	Sequence 5, Appl
38	54	15.6	392	4	US-09-996-243-205	Sequence 205, App
39	54	15.6	393	4	US-09-482-273-154	Sequence 154, App
40	54	15.6	878	4	US-09-463-238-4	Sequence 4, Appl
41	53.5	15.6	883	4	US-09-463-238-19	Sequence 19, Appl
42	53.5	15.5	396	4	US-09-198-452A-147	Sequence 147, App
43	53.5	15.5	632	4	US-09-252-991A-26091	Sequence 26091, A
44	53	15.3	312	3	US-09-216-295-21	Sequence 21, Appl
45	53	15.3	371	3	US-09-104-308-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-315-689-5  
Sequence 5, Application US/09315689  
Patent No. 6346510  
GENERAL INFORMATION:  
APPLICANT: Folman, Judah  
APPLICANT: O'Reilly, Michael  
TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions  
FILE REFERENCE: 05213-0229  
CURRENT APPLICATION NUMBER: US/09/315,689  
CURRENT FILING DATE: 1999-05-20  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 178  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-315-689-5

Query Match 100.0%; Score 346; DB 4; Length 178;  
Best Local Similarity 100.0%; Pred. No. 3e-38;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSPSWALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSVHGDSPN 60  
DB 60 ADRAAVPIVNLKDELLFSPSWALFSGSGEPLKPGARIFSPDGKDVLRHPTWPKSVHGDSPN 119  
QY 61 DPN 63  
DB 120 DPN 122

RESULT 2  
US-09-561-500-14  
Sequence 14, Application US/09561500  
Patent No. 6346219  
GENERAL INFORMATION:  
APPLICANT: Phillip E. Thorpe  
APPLICANT: Rolf A. Brecken  
TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF  
FILE REFERENCE: 4001.002500  
CURRENT APPLICATION NUMBER: US/09/561,500  
CURRENT FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/131,432  
PRIOR FILING DATE: 1999-04-28  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 14  
LENGTH: 182  
TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

OTHER INFORMATION: PEPTIDE

US-09-561-500-14

Query Match 100.0%; Score 346; DB 4; Length 182;

Best Local Similarity 100.0%; Pred. No. 3.1e-38;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSGDKDVLRRHPTWPKSVWHS 60

DB 64 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSGDKDVLRRHPTWPKSVWHS 123

QY 61 DPN 63

DB 124 DPN 126

## RESULT 3

US-09-561-108-14

Sequence 14, Application US/09561108

Patent No. 6342221

GENERAL INFORMATION:

APPLICANT: Philip E. Thorpe

APPLICANT: Rolf A. Brekken

TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF

FILE REFERENCE: 4001.002584

CURRENT APPLICATION NUMBER: US/09/561,108

PRIOR FILING DATE: 2000-04-28

PRIOR FILING DATE: 60/131,432

NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 14

LENGTH: 182

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

OTHER INFORMATION: PEPTIDE

US-09-561-108-14

Query Match 100.0%; Score 346; DB 4; Length 182;

Best Local Similarity 100.0%; Pred. No. 3.1e-38;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSGDKDVLRRHPTWPKSVWHS 60

DB 64 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSGDKDVLRRHPTWPKSVWHS 123

QY 61 DPN 63

DB 124 DPN 126

## RESULT 4

US-09-315-689-3

Sequence 3, Application US/09315689

Patent No. 6346510

GENERAL INFORMATION:

APPLICANT: Folkman, Judah

APPLICANT: O'Reilly, Michael

TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions

FILE REFERENCE: 05213-0229

CURRENT APPLICATION NUMBER: US/09/315,689

PRIOR FILING DATE: 1999-05-20

NUMBER OF SEQ ID NOS: 6

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3

LENGTH: 182

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-09-315-689-3

Query Match 100.0%; Score 346; DB 4; Length 182;

Best Local Similarity 100.0%; Pred. No. 3.1e-38;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSGDKDVLRRHPTWPKSVWHS 60

DB 64 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSGDKDVLRRHPTWPKSVWHS 123

QY 61 DPN 63

DB 124 DPN 126

## RESULT 5

US-09-561-526-14

Sequence 14, Application US/09561526

Patent No. 6416758

GENERAL INFORMATION:

APPLICANT: Philip E. Thorpe

APPLICANT: Rolf A. Brekken

TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF

FILE REFERENCE: 4001.002586

CURRENT APPLICATION NUMBER: US/09/561,526

PRIOR FILING DATE: 2000-04-28

PRIOR FILING DATE: 60/131,432

NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 14

LENGTH: 182

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

OTHER INFORMATION: PEPTIDE

US-09-561-526-14

Query Match 100.0%; Score 346; DB 4; Length 182;

Best Local Similarity 100.0%; Pred. No. 3.1e-38;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSGDKDVLRRHPTWPKSVWHS 60

DB 64 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSGDKDVLRRHPTWPKSVWHS 123

QY 61 DPN 63

DB 124 DPN 126

## RESULT 6

US-09-561-499-14

Sequence 14, Application US/09561499

Patent No. 6524583

GENERAL INFORMATION:

APPLICANT: Philip E. Thorpe

APPLICANT: Rolf A. Brekken

TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF

FILE REFERENCE: 4001.002582

CURRENT APPLICATION NUMBER: US/09/561,499

PRIOR FILING DATE: 2000-04-28

PRIOR FILING DATE: 60/131,432

NUMBER OF SEQ ID NOS: 44

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 14

LENGTH: 182

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC



OTHER INFORMATION: PEPTIDE  
US-09-561-499-14

Query Match 100.0%; Score 346; DB 4; Length 182;  
Best Local Similarity 100.0%; Pred. No. 3,1e-38;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
|||  
DB 64 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSPDGKDVLRHPTWPKSVWHS 123

QY 61 DPN 63  
|||  
DB 124 DPN 126

RESULT 7  
US-09-206-059-2  
Sequence 2, Application US/09206059  
Patent No. 6201104

GENERAL INFORMATION:  
APPLICANT: Macdonald, Nicholas  
APPLICANT: Sim, Kim Lee  
TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and  
FILE REFERENCE: 05213-0370  
CURRENT APPLICATION NUMBER: US/09/206,059  
CURRENT FILING DATE: 1998-12-04  
NUMBER OF SEQ ID NOS: 80  
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 2  
LENGTH: 183  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-206-059-2

Query Match 100.0%; Score 346; DB 3; Length 183;  
Best Local Similarity 100.0%; Pred. No. 3,1e-38;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
|||  
DB 64 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSPDGKDVLRHPTWPKSVWHS 123

QY 61 DPN 63  
|||  
DB 124 DPN 126

RESULT 8  
US-09-561-500-13  
Sequence 13, Application US/09561500  
Patent No. 6342219

GENERAL INFORMATION:  
APPLICANT: Philip E. Thorpe  
APPLICANT: Rolf A. Brekken  
TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF  
FILE REFERENCE: 4001.002500  
CURRENT APPLICATION NUMBER: US/09/561,500  
CURRENT FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/131,432  
PRIOR FILING DATE: 1999-04-28  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 13  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Artificial Sequence

FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-561-500-13  
Query Match 85.0%; Score 294; DB 4; Length 191;

Best Local Similarity 81.0%; Pred. No. 2,6e-31;  
Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
|||  
DB 71 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSPDGKDVLRHPTWPKSVWHS 130

QY 61 DPN 63  
|||  
DB 131 DPS 133

RESULT 9  
US-09-561-108-13  
Sequence 13, Application US/09561108  
Patent No. 6342221

GENERAL INFORMATION:  
APPLICANT: Philip E. Thorpe  
APPLICANT: Rolf A. Brekken  
TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF  
FILE REFERENCE: 4001.002584  
CURRENT APPLICATION NUMBER: US/09/561,108  
CURRENT FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/131,432  
PRIOR FILING DATE: 1999-04-28  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 13  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-561-108-13

Query Match 85.0%; Score 294; DB 4; Length 191;  
Best Local Similarity 81.0%; Pred. No. 2,6e-31;  
Matches 51; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSPDGKDVLRHPTWPKSVWHS 60  
|||  
DB 71 ADRAAVPIVNLKDELLFPGSEGLPKPGARIFSPDGKDVLRHPTWPKSVWHS 130

QY 61 DPN 63  
|||  
DB 131 DPS 133

RESULT 10  
US-09-561-526-13  
Sequence 13, Application US/09561526  
Patent No. 6416758

GENERAL INFORMATION:  
APPLICANT: Philip E. Thorpe  
APPLICANT: Rolf A. Brekken  
TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF  
FILE REFERENCE: 4001.002586  
CURRENT APPLICATION NUMBER: US/09/561,526  
CURRENT FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/131,432  
PRIOR FILING DATE: 1999-04-28  
NUMBER OF SEQ ID NOS: 44  
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 13  
LENGTH: 191  
TYPE: PRT  
ORGANISM: Artificial Sequence

FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-561-526-13  
Query Match 85.0%; Score 294; DB 4; Length 191;  
Best Local Similarity 81.0%; Pred. No. 2,6e-31;



Matches	49; Conservative	10; Mismatches	4; Indels	1; Gaps	1
Oy	1	ADBAAPVIV-NKDELTFPSWEALFSGSGEPLKPARIFPSFGKIVLHPTTPQKSVHMG	59		
Db	64	ADRGVPIVQNIKRDVLSFSPWDSLFGSQQLQPKARIFSGKIVLHPTTPQKSVHMG	123		
Oy	60	SDPN	63		
Db	124	SDPS	127		

RESULT 14  
US-08-159

Sequence 3, Application US/08159784  
Patent No. 5643783  
GENERAL INFORMATION:  
APPLICANT: Bjorn R. Olsen  
TITLE OF INVENTION: NOVEL COLLAGEN AND  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PC/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (version 5.0)  
SOFTWARE: Wordperfect (version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/159,784  
FILING DATE: December 1, 1993  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: John F. Freeman  
REGISTRATION NUMBER: 29,066  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191  
TYPE: amino acid  
STRANDEDNESS: N/A  
TOPOLOGY: N/A  
US-08-159-784-3

Query Match	54.9%	Score 190	DB 1	Length 191
Best Local Similarity	49.2%	Pred. No. 1.6e-17		
Matches 31	Conservative 16	Mismatches 16	Indels 0	Gaps 0

```
OY      I ADRAVPIVLKDELLFPSWEALFFSSEGPLKGARFPFDGQDVLHPPTWPKSVHWS   60
        |:::||||| :: ||::||| |::||| :||| :|||
Db       71 AERSLPIVLKGVLFNNWDISFGHGQGPPMHIPIVSPDRDIMDPSPMQKIWHGS    130
```

QY	61	DPN	63
		:	
Db	131	SPH	133

RESULT 15  
US-09-385-442-32  
; Sequence 32, Application US/09385442  
; Patent No. 6200954  
; GENERAL INFORMATION:  
; APPLICANT: Ge, Ruozen

```

APPLICANT: Kini, R. Manjunatha
TITLE OF INVENTION: Snail Peptides Having Potent Anti-Angiogenic Activity
FILE REFERENCE: 1781-170D
CURRENT APPLICATION NUMBER: US/09/385,442
CURRENT FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 60/099,313
EARLIER FILING DATE: 1999-09-04
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 32
LENGTH: 16
TYPE: PRT
ORGANISM: mammalian
FEATURES:
OTHER INFORMATION: Endo-4
US-09-385-442-32
Query Match      27.2%   Score 94; DB 3; Length 16;
Best Local Similarity 87.5%; Pred. No. 4e-06;
Matches    14; Conservative    1; Mismatches    1; Indels    0; Gaps    0;
QY          48 HPTWPKSVWHGSDPN 63
              ||| ||||| ||||| :
Db           1 HPATWPKSVWHGSDPS 16

```

Search completed: September 10, 2003, 14:13:05  
Job time : 22 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 14:09:48 ; Search time 41 Seconds  
(without alignments)  
243.897 Million cell updates/sec

Title: US-09-822-540A-1

Perfect score: 346  
Sequence: 1 ADRAAVPTVNLKDELLPFSW.....DVLRRPTWPKSVHMGSDPN 63

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT.\*
- 23: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT.\*
- 24: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	346	100.0	63	ABB08121	Human endostatin p
2	346	100.0	178	AAV94324	Alternate human en
3	346	100.0	178	AAU00900	Human Endostatin(T
4	346	100.0	179	AAU00901	Human Endostatin(T
5	346	100.0	180	AAU00899	Human Endostatin(T
6	346	100.0	181	AAU00898	Human Endostatin(T
7	346	100.0	182	AAV28399	Human endostatin.
8	346	100.0	182	AAV94323	Human endostatin p
9	346	100.0	182	AAV59622	Human endostatin p

10	346	100.0	182	AAU00897	Human Endostatin(T
11	346	100.0	182	AAU77951	Amino acid sequenc
12	346	100.0	183	AAV08693	Human endostatin p
13	346	100.0	183	AAU02113	SEQ ID 76 of W0991
14	346	100.0	183	AAV04993	Amino acid sequenc
15	346	100.0	183	AAV16451	Human endostatin p
16	346	100.0	183	AAV90771	Human angiotensin
17	346	100.0	183	AAV70252	Human angiotensin
18	346	100.0	183	AAU00896	Human Endostatin(T
19	346	100.0	183	AAV43379	Human endostatin S
20	346	100.0	183	AAV79901	Human endostatin p
21	346	100.0	183	AAV71332	Human endostatin.
22	346	100.0	183	AAV49503	Human endostatin p
23	346	100.0	183	AAV48895	Human endostatin p
24	346	100.0	183	AAV79753	Human endostatin.
25	346	100.0	184	AAV78717	Human vascular end
26	346	100.0	184	ABG31794	Murine endostatin
27	346	100.0	193	AAV90877	Human HMW endostat
28	346	100.0	193	AAV76690	Synthetic plasmid
29	346	100.0	195	AAV90874	Human HMW endostat
30	346	100.0	216	AAV30495	Amino acid sequenc
31	346	100.0	310	AAV76689	Synthetic plasmid
32	346	100.0	513	ABG73586	Human collagen XVI
33	346	100.0	682	ABP41878	Human Endostatin/I
34	346	100.0	682	AAV26327	Human ovarian anti
35	346	100.0	684	AAV26327	Human alpha-1 (VII
36	346	100.0	684	AAV25113	Human alpha-1 (VII
37	346	100.0	684	AAV17357	Human collagen typ
38	346	100.0	1301	AAV92296	Human alpha-1 (VII
39	346	100.0	1336	AAV08694	Human collagen 18
40	346	100.0	1336	AAV08694	Human endostatin p
41	346	100.0	1516	ABP66308	Human pancreatic c
42	346	100.0	1516	ABP66308	Human collagen XVI
43	346	98.8	183	AAV49810	Human endostatin p
44	335	96.8	271	AAV08407	A human angiotensin
45	302	87.3	184	AAV70265	Canine angiotensin

ALIGNMENTS

RESULT 1		
ID	ABB08121	Standard; protein; 63 AA.
AC	ABB08121;	
XX		
DT	10-SEP-2002	(first entry)
XX		
DE	Human endostatin polypeptide fragment JCK-362 (residues 52-114).	
XX		
KW	Endostatin; endothelial; cell proliferation; angiogenesis; cytostatic;	
KW	antithrombotic; antineuroleptic; ophthalmological; antidiabetic;	
KW	antitumor; vulnerrary; gynaecological; human.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Peptide	33..63
FT		/note= "JCK-367 peptide (ABB08122), specifically
FT	Modified-site	63
FT		claimed fragment"
XX		
PN	US2002058620-A1.	
XX		
PD	16-MAY-2002.	
XX		
PF	30-MAR-2001; 2001US-0822540.	
XX		
PR	03-APR-2000; 2000US-194561P.	
XX		
PA	(PHOE-) PHOENIX PHARMACOLOGIES INC.	

```

XX
PI Chang J;
XX
DR MPI; 2002-499519/53.
XX
PT New purified polypeptide useful for reducing tumor volume in an animal
XX PT complies a specified amino acid sequence
XX
PS Claim 1; Fig 1; 19pp; English.
XX
CC The invention relates to endostatin polypeptide fragments that can be
CC used as inhibitors of endothelial cell proliferation and angiogenesis.
CC The polypeptide inhibitors are useful for reducing tumor volume in an
CC animal; for inhibiting the proliferation of endothelial cells in an
CC animal; for targeting and binding purified polypeptide to an angiogenesis
CC receptor; inhibiting metastasis; for identifying, isolating or purifying
CC receptor proteins. They are also useful in therapy for treating diseases
CC such as atherosclerosis, haemangioma, leukemia, myocardial angiogenesis,
CC neovascular glaucoma, diabetic retinopathy, retrolental fibroplasia,
CC arthritis, diabetic neovascularisation, macular degeneration, wound
CC healing, peptic ulcer, fracture, keloid, hematopoiesis, ovulation,
CC menestruation and placentalion in human; and for producing monoclonal and
CC polyclonal antibodies. The present sequence represents a human endostatin
CC polypeptide fragment, JCR-352.
XX
SQ Sequence 63 AA;
XX
Query Match 100.0%; Score 346; DB 23; Length 63;
Best Local Similarity 100.0%; Pred. No. 3, 2e-39;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 ADRAAVPIVNLKDELLFPSWEALFSGSGEPKPKGARIFFDCKDVLRRHPTWPKSWMHGS 60
DB 1 ADRAAVPIVNLKDELLFPSWEALFSGSGEPKPKGARIFFDCKDVLRRHPTWPKSWMHGS 60
XX
QY 61 DPN 63
DB 61 DPN 63
XX
RESULT 2
AA94324
ID AAY94324 standard; Protein; 178 AA.
XX
AC AAY94324;
XX
DE 11-AUG-2000 (first entry)
XX
DE Alternate human endostatin protein.
XX
KW Human; endothelial cell proliferation inhibitor; collagen XVIII;
KW angiogenesis inhibitor; anti-tumour; cytostatic; antiproliferative;
KW vasotrophic; dermatological; ophthalmological; vulnary;
KW antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;
KW ocular angiogenic disease; atherosclerosis; scleroderma;
KW myocardial angiogenesis; telangiectasia; angiodiroma;
KW wound granulation.
XX
OS Homo sapiens.
XX
PN WO200026368-A2.
XX
PD 11-MAY-2000.
XX
PP 01-NOV-1999; 99WO-US25605.
XX
PR 30-OCT-1998; 98US-0106343.
XX PR 20-MAY-1999; 99US-0315689.
XX
PA (CHIL-) CHILDRENS MEDICAL CENT.
XX O'Reilly MS, Folkman MJ;
XX

```

```

DR MPI; 2000-365617/31.
XX
DR N-PSDB; AAA27005.
XX
PT Novel endostatin capable of inhibiting endothelial cell proliferation
XX PT and angiogenesis, useful for treating angiogenesis-dependent cancers
XX PT and as birth control agents
XX
PS Claim 3; Page 39; 68pp; English.
XX
CC The present sequence is an alternate functional endostatin
CC protein. When the human endostatin gene sequence AAA27004 is
CC recombinantly expressed, an observable doublet of protein results, both
CC versions of which are functional endostatin proteins. The present
CC endostatin variant is the same as the protein encoded by AAA27004 minus
CC the first four amino acids. Recombinant mouse endostatin (20 mg/kg) was
CC administered subcutaneously to mice implanted with Lewis lung carcinomas.
CC There was tumor mass regression non-detectable levels after 12 days of
CC therapy due to the angiogenesis inhibitory activity of endostatin. Thus
CC the protein is useful for treatment of angiogenesis-dependent cancers.
CC The polynucleotide and polypeptide sequences of this endostatin are
CC useful for treating and diagnosis of tumors, ocular angiogenic
CC diseases, Osler-Webber syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiodiroma
CC and wound granulation, for treatment of diseases related to excessive or
CC abnormal stimulation of endothelial cells e.g. intestinal adhesions,
CC atherosclerosis, scleroderma. The protein may also be useful as a birth
CC control agent by reducing or preventing uterine vascularisation. The
CC gene for endostatin may be isolated from cells or tissue that express
CC high levels of endostatin, eg. tumour cells, by generating cDNA from
CC mRNA using reverse transcriptase and then amplifying the DNA sequence.
XX
SQ Sequence 178 AA;
XX
Query Match 100.0%; Score 346; DB 21; Length 178;
Best Local Similarity 100.0%; Pred. No. 1, 2e-38;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 ADRAAVPIVNLKDELLFPSWEALFSGSGEPKPKGARIFFDCKDVLRRHPTWPKSWMHGS 60
DB 60 ADRAAVPIVNLKDELLFPSWEALFSGSGEPKPKGARIFFDCKDVLRRHPTWPKSWMHGS 119
XX
QY 61 DPN 63
DB 120 DPN 122
XX
RESULT 3
AAU00900
ID AAU00900 standard; Protein; 178 AA.
XX
AC AAU00900;
XX
DE 04-JUL-2001 (first entry)
XX
DE Human Endostatin(TM) N-terminal deletion mutant protein#2.
XX
KW Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubeosis;
KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW retinopathy of prematurity; macular corneal graft rejection;
KW neovascular glaucoma; retrolental fibroplasia; Osler-Webber Syndrome;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW haemophilic joint; angiodiroma; wound granulation; variant;
KW mutant; mutcin.
XX
OS Homo sapiens.
XX
PN WO200119989-A2.
XX
PD 22-MAR-2001.
XX

```

14-SEP-2000; 2000OWO-US25166.  
XX  
PR 14-SEP-1999; 99US-0153698.  
XX  
PA (ENTR-) ENTREMED INC.  
PI Liang H, Sim XL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ,  
PI Betrejo LL, Mierry FR, Shepard SR, Schrimsher JL,  
XX N-PSDB; AA00868.  
DR WP1: 2001-244802/25.  
XX  
PT Producing Endostatin protein for treating angiogenesis mediated  
PT diseases such as solid tumours, comprises recombinantly producing the  
PT protein using an expression system, and recovering and purifying the  
PT protein -  
XX  
Claim 5; Page 33; 67pp; English.

The sequence represents Human Endostatin(TM) N-terminal deletion mutant protein lacking the N-terminal 4 amino acids and the C-terminal lysine, a natural variant recovered from fermentations of *Pichia pastoris* cultures harbouring a expression plasmid containing the Endostatin(TM) DNA sequence given in AA00868. The new method of the invention is useful for producing, recovering and purifying Endostatin(TM) from biological sources, such as biological fluids, tissues, cells, culture media, and fermentation media. Endostatin(TM) is useful for treating angiogenesis mediated diseases such as solid tumours, blood borne tumours, leukemias, tumour metastases, benign tumours, e.g. haemangioma, acoustic neuroma, neurofibromas, trachomas, and pyogenic granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases, e.g., diabetic retinopathy, retinopathy of prematurity, macular degeneration, corneal graft rejection, neovascular glaucoma, colon cancer, retrorenal fibroplasia, rubecosis, Osler-Webber Syndrome, myocardial angiogenesis, plaque neovascularisation, telangiectasia, haemophilic joints, angiofibroma, and wound granulation. Endostatin(TM) is also useful for treating disease of excessive or abnormal stimulation of endothelial cells such as intestinal adhesions, atherosclerosis, scleroderma and hypertrophic scars. Higher yields of more purified, and biologically active Endostatin(TM) are obtained by the new method. Endostatin(TM) can be stored in buffers for extended periods of time, and also subjected to lyophilisation, while preserving biological activity. Centrifugation of broth from fermentation steps in production is avoided, preventing unwanted potential cellular lysis and contamination with additional proteins, pigments, enzymes and other cellular chemicals and debris.

Sequence 178 AA;

	Query Match	Best Local Similarity	Score 346;	DB 22;	Length 178;
	Matches 63;	Conservative 100.0%;	Fred. No. 1.2e-38;	Mismatches 0;	Indels 0;
				Gaps 0	
OY	1 ADRAAVPINLKDDELPSEWALFSGSEPEIKRGARIFPFDDGKDVLRHPTPKSVWRGS 60				
Db	60 ABRAAVPINLKDELFPSEWALFSGSDEPLRKGARIFSFDDGKDVLRHPTPKSVWRHS 119				
OY	61 DFN 63				
Db	120 DFN 122				

result 4  
AAU00901  
ID AAU00901 standard; Protein; 179 AA.  
AC AAU00901;  
XX  
XX DT 04-JUL-2001 (first entry)  
XX  
DE Human Endostatin(TM) N-terminal mutant protein#1.  
KW Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;

OS	Homo sapiens.
XX	
PN	WO200119989-A2.
XX	
PD	22-MAR-2001.
XX	
PF	14-SEP-2000; 2000WO-US25166.
XX	
FR	14-SEP-1999; 99US-0153696.
PA	(ENTR-) ENTREMED INC.
PI	Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ,
DR	Bernabeo JL, Mistery FR, Shepard SR, Schrimsher JL;
XX	WPI: 2001-244802/25.
N-PDB:	AAS00868.
PT	Producing Endostatin protein for treating angiogenesis mediated
PT	diseases such as solid tumours, comprises recombinantly producing the
PT	protein using an expression system, and recovering and purifying the
PS	Claim 5; Page 32; 67pp; English.
XX	
CC	The sequence represents a Human Endostatin(TM) N-terminal deletion
CC	mutant lacking the N-terminal 4 amino acids. The new method of the
CC	invention is useful for producing, recovering and purifying Endostatin
CC	(TM) from biological sources, such as biological fluids, tissues, cells,
CC	culture media, and fermentation media. Endostatin(TM) is useful for
CC	treating angiogenesis mediated diseases such as solid tumours, blood
CC	borne tumours, leukemias, tumour metastases, benign tumours, e.g.
CC	haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic
CC	granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,
CC	e.g., diabetic retinopathy, retinopathy of prematurity, macular
CC	degeneration, corneal graft rejection, neovascular glaucoma, colon
CC	cancer, retroileal fibroplasia, rubecosis, Osler-Webber Syndrome,
CC	myocardial angiodysplasia, plaque neovascularisation, telangiectasia,
CC	haemophilic joints, angiofibroma, and wound granulation. Endostatin(TM)
CC	is also useful for treating disease of excessive or abnormal stimulation
CC	of endothelial cells such as intestinal adhesions, atherosclerosis, and
CC	scleroderma and hypertrophic scars. Higher yields of more purified, and
CC	biologically active Endostatin(TM) are obtained by the new method.
CC	Endostatin(TM) can be stored in buffers for extended periods of time, and
CC	also subjected to lyophilisation, while preserving biological activity.
CC	Centrifugation of broth from fermentation steps in production is avoided,
CC	preventing unwanted potentially cellular lysis and contamination with
CC	additional proteins, pigments, enzymes and other cellular chemicals and
CC	debris.
SQ	Sequence 179 AA:
XX	
Query Match	100.0%; Score 346; DB 22; Length 179;
Best Local Similarity	100.0%; Pred. No. 1,2e-38;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	1 ADRAAVPTNLTDELLFSPWEALFGSGSEPLPKGARIPEFDGKVLRHPTPOKSWMHS 60
DB	
DB	60 ADRAAVPTNLTDELLFSPWEALFGSGSEPLPKGARIPEFDGKVLRHPTPOKSWMHS 119
QY	61 DPN 63
DB	
DB	120 DPN 122

RESULT 5  
AAU00899  
ID AAU00899 standard; Protein: 180 AA.  
XX  
AC AAU00899;  
XX  
DT 04-JUL-2001 (first entry)  
XX  
DE Human Endostatin(TM) C-terminus minus 3 protein.  
XX  
KW Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;  
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;  
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubecosis;  
KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;  
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
KW retinopathy of prematurity; macular corneal graft rejection;  
KW neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;  
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
KW haemophilic joint; angiofibroma; wound granulation; variant;  
KW C-terminus minus 3 protein.  
XX  
OS Homo sapiens.  
XX  
PN WO200119989-A2.  
XX  
PD 22-MAR-2001.  
XX  
PF 14-SEP-2000; 2000WO-US25166.  
XX  
PR 14-SEP-1999; 99US-0153698.  
XX  
PA (ENTR-) ENTREMED INC.  
XX  
PI Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;  
PI Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;  
XX  
DR WPI; 2001-244802/25.  
XX  
DR N-PSDB; AAS00867.  
XX  
PT Producing Endostatin protein for treating angiogenesis mediated  
PT diseases such as solid tumours, comprises recombinantly producing the  
PT protein using an expression system, and recovering and purifying the  
PT protein.  
XX  
PS Claim 5; Page 30; 67pp; English.  
XX  
CC The sequence represents Human Endostatin(TM) C-terminus minus 3  
CC protein, a natural variant lacking the C-terminal 3 amino acids of  
CC Endostatin(TM) recovered from fermentations of Pichia  
CC pastoris cultures harbouring a expression plasmid containing the  
CC Endostatin(TM) DNA sequence given in AAS00867. The new method of the  
CC invention is useful for producing, recovering and purifying Endostatin  
CC (TM) from biological sources, such as biological fluids, tissues, cells,  
CC culture media, and fermentation media. Endostatin(TM) is useful for  
CC treating angiogenesis mediated diseases such as solid tumours, blood  
CC borne tumours, leukaemias, tumour metastases, benign tumours, e.g.  
CC haemangioma, acoustic neuroma, neurofibromas, trachomas, and pyogenic  
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,  
CC e.g., diabetic retinopathy, retinopathy of prematurity, macular  
CC degeneration, corneal graft rejection, neovascular glaucoma, colon  
CC cancer, retrolental fibroplasia, rubecosis, Osler-Weber Syndrome,  
CC myocardial angiogenesis, plaque neovascularisation, telangiectasia,  
CC is also useful for treating disease of excessive or abnormal stimulation  
CC of endothelial cells such as intestinal adhesions, atherosclerosis,  
CC scleroderma and hypertrophic scars. Higher yields of more purified, and  
CC biologically active Endostatin(TM) are obtained by the new method.  
CC Endostatin(TM) can be stored in buffers for extended periods of time, and  
CC also subjected to lyophilisation, while preserving biological activity.  
CC Centrifugation of broth from fermentation steps in production is avoided,  
CC preventing unwanted potential cellular lysis and contamination with  
CC additional proteins, pigments, enzymes and other cellular chemicals and

CC debris.  
XX  
SQ Sequence 180 AA;  
XX  
Query Match 100.0%; Score 346; DB 22; Length 180;  
Best Local Similarity 100.0%; Pred. No. 1-2e-39;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 ADRAAVPIVNLKDELLFSGSEGLKPGARIFSGDKDVLRRPTPQKSWHGS 60  
DB 64 ADRAAVPIVNLKDELLFSGSEGLKPGARIFSGDKDVLRRPTPQKSWHGS 123  
QY 61 DPN 63  
DB 124 DPN 126  
XX  
RESULT 6  
AAU00898  
ID AAU00898 standard; Protein: 181 AA.  
XX  
AC AAU00898;  
XX  
DT 04-JUL-2001 (first entry)  
XX  
DE Human Endostatin(TM) C-terminus minus 2 protein.  
XX  
KW Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;  
KW blood borne tumour; leukaemia; tumour metastasis; benign tumour;  
KW haemangioma; acoustic neuroma; neurofibroma; trachoma; rubecosis;  
KW pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;  
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
KW retinopathy of prematurity; macular corneal graft rejection;  
KW neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;  
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
KW haemophilic joint; angiofibroma; wound granulation; variant;  
KW C-terminus minus 2 protein.  
XX  
OS Homo sapiens.  
XX  
PN WO200119989-A2.  
XX  
PD 22-MAR-2001.  
XX  
PF 14-SEP-2000; 2000WO-US25166.  
XX  
PR 14-SEP-1999; 99US-0153698.  
XX  
PA (ENTR-) ENTREMED INC.  
XX  
PI Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;  
PI Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;  
XX  
DR WPI; 2001-244802/25.  
XX  
DR N-PSDB; AAS00867.  
XX  
PT Producing Endostatin protein for treating angiogenesis mediated  
PT diseases such as solid tumours, comprises recombinantly producing the  
PT protein using an expression system, and recovering and purifying the  
PT protein.  
XX  
PS Claim 5; Page 30; 67pp; English.  
XX  
CC The sequence represents Human Endostatin(TM) C-terminus minus 2  
CC protein, a natural variant lacking the C-terminal 2 amino acids of  
CC Endostatin(TM) recovered from fermentations of Pichia  
CC pastoris cultures harbouring a expression plasmid containing the  
CC Endostatin(TM) DNA sequence given in AAS00867. The new method of the  
CC invention is useful for producing, recovering and purifying Endostatin  
CC (TM) from biological sources, such as biological fluids, tissues, cells,  
CC culture media, and fermentation media. Endostatin(TM) is useful for  
CC treating angiogenesis mediated diseases such as solid tumours, blood  
CC borne tumours, leukaemias, tumour metastases, benign tumours, e.g.



CC haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic  
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,  
 CC e.g., diabetic retinopathy, retinopathy of prematurity, macular  
 CC degeneration, corneal graft rejection, neovascular glaucoma, colon  
 CC cancer, retrorenal fibroplasia, rubeosis, Osler-Weber Syndrome,  
 CC myocardial angiogenesis, plaque neovascularisation, telangiectasia,  
 CC haemophilic joints, angiodioma, and wound granulation. Endostatin(TM)  
 CC is also useful for treating disease of excessive or abnormal stimulation  
 CC of endothelial cells such as intestinal adhesions, atherosclerosis,  
 CC scleroderma and hypertrophic scars. Higher yields of more purified, and  
 CC biologically active Endostatin(TM) are obtained by the new method.  
 CC Endostatin(TM) can be stored in buffers for extended periods of time, and  
 CC also subjected to lyophilisation, while preserving biological activity.  
 CC Centrifugation of broth from fermentation steps in production is avoided,  
 CC preventing unwanted potential cellular lysis and contamination with  
 CC additional proteins, pigments, enzymes and other cellular chemicals and  
 CC debris.

CC Sequence 181 AA;

Query Match 100.0%; Score 346; DB 22; Length 181;

Best Local Similarity 100.0%; Pred. No. 1.2e-38;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSEWALFSGSGPLKPGARIFSPDGKDVLRHPTWPKSVWHGS 60  
 DB 64 ADRAAVPIVNLKDELFPSEWALFSGSGPLKPGARIFSPDGKDVLRHPTWPKSVWHGS 123

QY 61 DPN 63  
 DB 124 DPN 126

RESULT 7  
 ID AAB28399 standard; Protein; 182 AA.

AC AAB28399;

DT 19-FEB-2001 (first entry)

DE Human endostatin.

XX Human; endostatin; cytosolic; antiproliferative;

KM vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;

XX cancer; vascularised solid tumour.

OS Homo sapiens.

XX WO200064946-A2.

XX 02-NOV-2000.

PF 28-APR-2000; 2000WO-US11367.

XX 28-APR-1999; 99US-0131432.

PA (TEXA) UNIV TEXAS SYSTEM.

PI Thorpe PE, Brecken RA,

XX WPI; 2000-687317/67.

PT Immunogenic composition for the treatment and diagnosis of cancer

CC comprises an anti-VEGF (vascular endothelial growth factor) antibody

CC binding the same epitope as the monoclonal antibody ATCC PTA 1595 -

XX Example 10; Page 291-292; 298pp; English.

XX The present invention relates to anti-Vascular Endothelial Growth Factor

CC (VEGF) antibodies that bind to the same epitope as the monoclonal

CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to

CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF

CC receptor VEGFR1. The present sequence is human endostatin. Endostatin  
 CC may be conjugated onto the anti-VEGF antibodies of the present invention.  
 CC The anti-VEGF antibodies of the present invention are useful for the  
 CC treatment and diagnosis of cancer, especially vascularised solid tumours.

CC Sequence 182 AA;

Query Match 100.0%; Score 346; DB 21; Length 182;

Best Local Similarity 100.0%; Pred. No. 1.2e-38;

Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELFPSEWALFSGSGPLKPGARIFSPDGKDVLRHPTWPKSVWHGS 60  
 DB 64 ADRAAVPIVNLKDELFPSEWALFSGSGPLKPGARIFSPDGKDVLRHPTWPKSVWHGS 123

QY 61 DPN 63  
 DB 124 DPN 126

RESULT 8

AA94323  
 ID AA94323 standard; Protein; 182 AA.

AC AA94323;

DT 11-AUG-2000 (first entry)

DE Human endostatin protein.

XX Human; endostatin cell proliferation inhibitor; collagen XVII;

KM angiogenesis inhibitor; anti-tumour; cytosolic; antiproliferative;

KM vasotrophic; dermatological; ophthalmological; vulvexary;

KM antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;

KM ocular angiogenic disease; atherosclerosis; scleroderma;

KM myocardial angiogenesis; telangiectasia; angiodioma;

XX wound granulation.

OS Homo sapiens.

XX WO200026368-A2.

XX 11-MAY-2000.

PF 01-NOV-1999; 99WO-US25605.

XX 30-OCT-1998; 98US-0106343.

PR 20-MAY-1999; 99US-0315689.

PA (CHIL-) CHILDRENS MEDICAL CENT.

PI O'Reilly MS, Folkman MJ;

XX WPI; 2000-365617/31.

DR N-PSDB; AA27004.

XX Novel endostatin capable of inhibiting endothelial cell proliferation

CC and angiogenesis, useful for treating angiogenesis-dependent cancers

CC as birth control agents

XX Claim 2; Page 38; 68pp; English.

XX The present sequence is an endostatin protein which is the carboxy

CC terminal protein of human collagen XVII. Recombinant mouse endostatin

CC (20 mg/kg) was administered subcutaneously to mice implanted with Lewis

CC lung carcinomas. There was tumour mass regression non-detectable levels

CC after 12 days of therapy due to the angiogenesis inhibitory activity of

CC endostatin. Thus the protein is useful for treatment of angiogenesis-

CC dependent cancers. The polynucleotide and polypeptide sequences of this

CC endostatin are useful for treating and diagnosis of tumours, ocular

CC angiogenic diseases, Osler-Weber syndrome, myocardial angiogenesis,

CC plaque neovascularisation, telangiectasia, haemophilic joints,

CC angiodioma and wound granulation, for treatment of diseases related to

CC excessive or abnormal stimulation of endothelial cells e.g. intestinal  
 CC adhesions, atherosclerosis, scleroderma. The protein may also be useful  
 CC as a birth control agent by reducing or preventing uterine  
 CC vascularization. The gene for endostatin may be isolated from cells or  
 CC tissue that express high levels of endostatin, eg. tumour cells, by  
 CC generating cDNA from mRNA using reverse transcriptase and then amplifying  
 CC the DNA sequence.

SO Sequence 182 AA;

Query Match 100.0%; Score 346; DB 21; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-38;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSPWEALFSGSEGLPKRGARIFSPDGKDVLRHPTWPKSVWHS 60  
 DB 64 ADRAAVPIVNLKDELLFSPWEALFSGSEGLPKRGARIFSPDGKDVLRHPTWPKSVWHS 123

QY 61 DPN 63  
 DB 124 DPN 126

RESULT 9  
 AAY59622

ID AAY59622 standard; protein; 182 AA.

AC AAY59622;

DT 14-MAR-2000 (first entry)

DE Human endostatin protein fragment.

XX Endostatin; scatter factor activity; human; tubulogenesis; psoriasis;  
 XX metastatic cancer; tumorigenesis; ocular angiogenic disease;  
 XX rheumatoid arthritis; Osler-Weber syndrome; telangiectasia;  
 XX haemophilic joint; angiodioma; wound granulation.

OS Homo sapiens.

PN WO962944-A2.

PD 09-DEC-1999.

PF 03-JUN-1999; 99WO-US12278.

PR 03-JUN-1998; 98US-0087890.

PR 10-JUL-1998; 98US-0092393.

PR 01-SEP-1998; 98US-0098790.

PA (CHIL-) CHILDRENS MEDICAL CENT.

PI Javaherian K, Folkman MJ;

DR WPI; 2000-072833/06.

PT New endostatin oligomers, used for treating e.g. tumours -

PS Disclosure; Page 6; 44pp; English.

CC This sequence is a fragment of the human endostatin protein. Endostatin  
 CC is an approximately 20kD C-terminal globular domain of the collagen-like  
 CC protein collagen XVIII. Protein oligomers consisting of more than one  
 CC endostatin monomer have anti-tubulogenic effects and induce  
 CC reorganization of the actin cytoskeleton. The oligomer has scatter factor  
 CC activity. The oligomers induce the destruction of tubular lumens and  
 CC elongation of cells, and inhibit tubulogenesis and tumorigenesis. The  
 CC oligomers can also be used to treat metastatic cancers, tumours,  
 CC rheumatoid arthritis, psoriasis, ocular angiogenic disease, Osler-Weber  
 CC syndrome, plaque neovascularisation, telangiectasia, haemophilic  
 CC joints, angiodioma and wound granulation. The oligomers can also be  
 CC used to treat diseases that have angiogenesis as a pathological  
 CC consequence e.g. ulcers. The endostatin oligomers can also be used to

CC develop affinity columns for isolating antibodies or receptors. Passive  
 CC antibody therapy using antibodies that specifically bind endostatin  
 CC oligomers can be used to modulate morphogenic processes such as  
 CC metastatic cancer as well as angiogenesis-dependent processes such as  
 CC reproduction, development, wound healing, tissue repair, and  
 CC angiogenesis-dependent diseases. Also, antisera directed to the Fab  
 CC regions of endostatin oligomer antibodies can be administered to block  
 CC the ability of endogenous endostatin oligomer antisera to bind endostatin  
 CC oligomers.

SO Sequence 182 AA;

Query Match 100.0%; Score 346; DB 21; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-38;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADRAAVPIVNLKDELLFSPWEALFSGSEGLPKRGARIFSPDGKDVLRHPTWPKSVWHS 60  
 DB 64 ADRAAVPIVNLKDELLFSPWEALFSGSEGLPKRGARIFSPDGKDVLRHPTWPKSVWHS 123

QY 61 DPN 63  
 DB 124 DPN 126

RESULT 10  
 AAU0897

ID AAU0897 standard; protein; 182 AA.

AC AAU0897;

DT 04-JUL-2001 (first entry)

DE Human Endostatin(TM) C-terminus minus 1 protein.

XX Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;  
 XX blood borne tumour; leukaemia; tumour metastasis; benign tumour;  
 XX haemangioma; acoustic neuroma; neurofibroma; trachoma; tubecosis;  
 XX pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;  
 XX ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
 XX neuropathy of prematurity; macular corneal graft rejection;  
 XX neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;  
 XX myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
 XX haemophilic joint; angiodioma; wound granulation; variant;  
 XX C-terminus minus 1 protein.

OS Homo sapiens.

PN WO200119989-A2.

PD 22-MAR-2001.

PF 14-SEP-2000; 2000WO-US25166.

PR 14-SEP-1999; 99US-0153698.

PA (ENTR-) ENTREMED INC.

PI Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;

XX Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;

DR WPI; 2001-244802/25.

PS N-PSDB; AAS00897.

PT Producing Endostatin protein for treating angiogenesis mediated  
 PT diseases such as solid tumours, comprises recombinantly producing the  
 PT protein using an expression system, and recovering and purifying the  
 PT protein -

PS Claim 5; Page 30; 67pp; English.

CC The sequence represents Human Endostatin(TM) C-terminus minus 1  
 CC protein, a natural variant lacking the C-terminal amino acid of

CC Endostatin(TM) recovered from fermentations of Pichia  
 CC pastoris cultures harbouring a expression plasmid containing the  
 CC Endostatin(TM) DNA sequence given in AAS00867. The new method of the  
 CC invention is useful for producing, recovering and purifying Endostatin  
 CC (TM) from biological sources, such as biological fluids, tissues, cells,  
 CC culture media, and fermentation media. Endostatin(TM) is useful for  
 CC treating angiogenesis mediated diseases such as solid tumours, blood  
 CC borne tumours, leukemias, tumour metastases, benign tumours, e.g.  
 CC haemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic  
 CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,  
 CC e.g., diabetic retinopathy, retinopathy of prematurity, macular  
 CC degeneration, corneal graft rejection, neovascular glaucoma, colon  
 CC cancer, retrolental fibroplasia, rubeosis, Osler-Weber Syndrome,  
 CC myocardial angiogenesis, plaque neovascularisation, telangiectasia,  
 CC haemophilic joints, angiodiroma, and wound granulation. Endostatin(TM)  
 CC is also useful for treating disease of excessive or abnormal stimulation  
 CC of endothelial cells such as intestinal adhesions, atherosclerosis,  
 CC scleroderma and hypertrophic scars. Higher yields of more purified, and  
 CC biologically active Endostatin(TM) are obtained by the new method.  
 CC Endostatin(TM) can be stored in buffers for extended periods of time, and  
 CC also subjected to lyophilisation, while preserving biological activity.  
 CC Centrifugation of broth from fermentation steps in production is avoided,  
 CC preventing unwanted potential cellular lysis and contamination with  
 CC additional proteins, pigments, enzymes and other cellular chemicals and  
 CC debris.

CC Sequence 182 AA:

Query Match 100.0%; Score 346; DB 22; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-38;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLKDELFPSEWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKQSVWHS 60  
 DB 64 ADRAAVPIVNLKDELFPSEWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKQSVWHS 123

OY 61 DPN 63  
 DB 124 DPN 126

RESULT 11  
 ID AAV77951 standard; Protein; 182 AA.

XX AC AAV77951;

DT 02-JUL-2002 (first entry)

DE Amino acid sequence for human endostatin.

XX Human; immunoglobulin; anti-vascular endothelial growth factor antibody;  
 XX anti-VEGF antibody; monoclonal antibody 2C3 ATCC PTA 1595; VEGF receptor;  
 XX VEGFR2; KDR/Flk-1; VEGFR1; Flt-1; angiogenesis; macular degeneration;  
 XX ocular neovascular disease; cancer; vascularised solid tumour; AIDS;  
 XX metastatic tumour; endothelial cell proliferation; inflammatory disorder;  
 XX atherosclerosis; diabetic retinopathy; corneal graft rejection;  
 XX acquired immune deficiency syndrome; infection; restenosis; fungal ulcer;  
 XX sickle cell anaemia; endometriosis; endostatin.

XX Homo sapiens.

XX AU200179401-A.

XX 06-DEC-2001.

XX 12-OCT-2001; 2001AU-0079401.

XX 28-APR-2000; 2000AU-0048049.

XX 12-OCT-2001; 2001AU-0079401.

XX (TEXA ) UNIV TEXAS SYSTEM.

PI Thorpe PE, Brekken RA;  
 XX WPI; 2002-281368/33.  
 XX Immunoglobulin compositions for treating cancer by inhibiting  
 XX angiogenesis and for delivering a diagnostic agent to tumour, comprises  
 XX anti-vascular endothelial growth factor antibody attached to a  
 XX biological agent

XX Example 10; Page 12-13 (Sequence listing); 300pp; English.

XX The present invention relates to antibody-based compositions comprising  
 XX an immunoglobulin such as anti-vascular endothelial growth factor  
 XX (VEGF) antibody (Ab) (or its antigen-binding fragment), attached to a  
 XX biological agent, where the Ab binds to the same epitope as the  
 XX monoclonal antibody (Mab) 2C3 ATCC PTA 1595, and significantly inhibits  
 XX VEGF binding to the VEGF receptor VEGFR2 (KDR/Flk-1) without inhibiting  
 XX VEGF binding to the VEGF receptor VEGFR1 (Flt-1). The compositions  
 XX of the invention are useful in therapy, and diagnosis, for inhibiting  
 XX angiogenesis in an animal having ocular neovascular disease or macular  
 XX degeneration, and for delivering a biological agent to a vascularised  
 XX tumour. The compositions can also be used for treating cancer and  
 XX subjects at risk of developing, a vascularised solid tumour, a metastatic  
 XX tumour or metastases from a primary tumour. The composition is useful  
 XX for specifically inhibiting VEGF-induced endothelial cell proliferation,  
 XX without significantly inhibiting VEGF-induced macrophage, osteoclast or  
 XX chondrocyte function. The compositions can be used for treating various  
 XX diseases such as inflammatory disorders, atherosclerosis, diabetic  
 XX retinopathy, restenosis, acquired immune deficiency syndrome (AIDS),  
 XX blood borne tumours, corneal graft rejection, Crohn's disease, fungal  
 XX ulcers, infections, sickle cell anaemia, and endometriosis. The present  
 XX sequence represents human endostatin. Endostatin may be attached or  
 XX functionally associated with anti-VEGF antibodies.

SO Sequence 182 AA:

Query Match 100.0%; Score 346; DB 23; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-38;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ADRAAVPIVNLKDELFPSEWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKQSVWHS 60  
 DB 64 ADRAAVPIVNLKDELFPSEWALFSGSEGPLKPGARIFSPDGKDYLRHPTWPKQSVWHS 123

OY 61 DPN 63

DB 124 DPN 126

RESULT 12

ID AAY08693 standard; Protein; 183 AA.

XX AC AAY08693;

DT 10-AUG-1999 (first entry)

DE Human endostatin protein fragment.

XX Plasmidogen; human; angiostatin; endostatin; gene therapy; vector;  
 XX anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;  
 XX tumour growth; solid tumour; diabetic retinopathy; retina.

XX Homo sapiens.

XX MO9926480-A1.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-US24950.

XX 20-NOV-1997; 97US-0975424.

PA (GENE-) GENETIX PHARM INC.  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 XX  
 PI Bachelot T, Leboulch P, Pawlinski R;  
 XX WPI, 1999-357696/30.  
 DR N-PSDB; AAX77719.  
 XX  
 PT Anti-angiogenic gene therapy vectors  
 XX  
 PS Disclosure; Page 74-75; 83pp; English.  
 XX  
 CC This invention describes a novel viral gene therapy vector comprising a  
 CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
 CC from human or murine endostatin, human or murine endostatin and  
 CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
 CC efficiently attenuated for use in human gene therapy. The products of  
 CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
 CC ophthalmological activity. The vector is used in gene therapy for  
 CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
 CC expresses an anti-angiogenic polypeptide, where the anti-angiogenic polypeptide  
 CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
 CC inhibits angiogenesis in the vicinity of the retina. The vector is  
 CC administered to cells ex vivo and then administered to the patient.  
 CC  
 SQ Sequence 183 AA;  
 XX  
 Query Match 100.0%; Score 346; DB 20; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-38;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ADRAAVPIVNLKDELFPSEWALFSGSEGLPKGARIFFDQKDVLRHPTWPKSVWHS 60  
 DB 64 ADRAAVPIVNLKDELFPSEWALFSGSEGLPKGARIFFDQKDVLRHPTWPKSVWHS 123  
 QY 61 DPN 63  
 DB 124 DPN 126  
 DB

RESULT 13  
 AAY02113  
 ID AAY02113 standard; Protein; 183 AA.  
 XX  
 AC AAY02113;  
 XX  
 DT 16-JUL-1999 (first entry)  
 XX  
 DE SEQ ID 76 of WO9916889.  
 XX  
 KM Angiostatin; endostatin; interferon; thrombospondin;  
 KM interferon-inducible protein; platelet factor 4; anti-angiogenic;  
 KM anti-tumor; multifunctional protein; angiogenic-mediated disease;  
 KM cancer; diabetic retinopathy; macular degeneration; arthritis;  
 KM tumor cell production.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9916889-A1.  
 XX  
 PD 08-APR-1999.  
 XX  
 PF 30-SEP-1998; 98WO-US20464.  
 XX  
 PR 01-OCT-1997; 97US-0060609.  
 XX  
 PA (SEAR ) SEARLE & CO G D.  
 XX  
 PI Bolanowski MA, Caparon MH, Casperson GF, Gregory SA;  
 PI Klein BK, McKearn JP;  
 XX WPI, 1999-255098/21.  
 XX

PT New multifunctional proteins useful for treating angiogenic-mediated  
 PT diseases  
 XX  
 XX  
 PS Disclosure; Page 106-107; 121pp; English.  
 XX  
 CC The specification describes multifunctional proteins which comprise  
 CC combinations of angiostatin, endostatin, interferon, thrombospondin,  
 CC interferon-inducible protein and platelet factor 4, and have  
 CC anti-angiogenic and/or anti-tumor activity. The multifunctional protein  
 CC may exhibit useful properties such as having similar or greater  
 CC biological activity when compared to a single factor or by having  
 CC improved half-life or decreased adverse side effects, or a combination  
 CC of these properties. The proteins can be used for treating an  
 CC angiogenic-mediated disease, e.g. cancer, diabetic retinopathy, macular  
 CC degeneration, or arthritis. They can also be used for inhibiting the  
 CC production of tumor cells (characteristic of lung, breast, ovarian,  
 CC prostate, pancreatic, gastric, colon, renal, bladder cancers; melanoma,  
 CC hepatoma, sarcoma and lymphoma) in a patient and for inhibiting tumor  
 CC growth. The present sequence is used in the course of the invention.  
 CC  
 SQ Sequence 183 AA;  
 XX  
 Query Match 100.0%; Score 346; DB 20; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-38;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ADRAAVPIVNLKDELFPSEWALFSGSEGLPKGARIFFDQKDVLRHPTWPKSVWHS 60  
 DB 64 ADRAAVPIVNLKDELFPSEWALFSGSEGLPKGARIFFDQKDVLRHPTWPKSVWHS 123  
 QY 61 DPN 63  
 DB 124 DPN 126  
 DB

RESULT 14  
 AAB30493  
 ID AAB30493 standard; Protein; 183 AA.  
 XX  
 AC AAB30493;  
 XX  
 DT 06-MAR-2001 (first entry)  
 XX  
 DE Amino acid sequence of human endostatin encoded by plasmid pMALCH15.  
 XX  
 KM Streptomyces sp. strain C5; SnpA; S. venezuelae; alpha-amylase;  
 KM endostatin; cancer; tumour growth; angiogenesis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2000060945-A1.  
 XX  
 PD 19-OCT-2000.  
 XX  
 PF 12-APR-2000; 2000WO-US09747.  
 XX  
 PR 13-APR-1999; 99US-0129084.  
 XX  
 PA (MERI ) MERCK & CO INC.  
 XX  
 PI Desautel CL, Strohl WR;  
 XX  
 DR WPI; 2000-686970/67.  
 XX  
 N-PSDB; AAC62023.  
 XX  
 PT Preparation of soluble recombinant endostatin involves transforming  
 PT Streptomyces host with expression vector comprising nucleotide  
 PT sequence encoding endostatin operably linked to linker and leader  
 PT peptide -  
 XX  
 PS Example 1; Fig 6; 57pp; English.  
 XX  
 CC The present sequence represents human endostatin. The protein is

expressed in *Streptomyces*. Leader sequences of *Streptomyces* sp. strain C5 SmpA and *S. venezuelae* alpha-amylase proteins are linked to the N-terminal of endostatin. This ensures that endostatin protein is produced as a secreted, soluble protein which needs no refolding, is stable in the fermentation broth and is produced in large quantities. The method is used for preparing soluble recombinant human, murine or primate endostatin, which is useful in the treatment of cancer, inhibition of tumour growth, inhibition of angiogenesis, isolation of receptors for endostatin and for identification of anti-angiogenic compounds in assays. The endostatin protein is produced as a secreted, soluble protein which needs no refolding, is stable in the fermentation broth and is produced in large quantities. *Streptomyces* are amenable for cultivation in large fermentations allowing for large quantities of soluble endostatin to be produced.

**SQ Sequence 183 AA;**

Query Match	100.0%	Score 346;	DB 21;	Length 183;
Best Local Similarity	100.0%	Pred. No. 1.2e-38;		
Matches 63; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

QY 1 ADRAAVPIVNLKDELLFSPSEALFSGEGPLKPGARIFSFDGNDVLRHPTWPQKSVHGS 60  
Db 64 ADRAAVPIVNLKDELLFSPSEALFSGEGPLKPGARIFSFDGNDVLRHPTWPQKSVHGS 123

QY	61	DPN	63
Db	124	DPN	126

## RESULT 1.5

ID AAB16451 standard; 183 AA

AC AAB16451;

DT 27-OCT-2000 (first entry)

Human endostatin protein sequence

KM Angiogenesis-inhibiting protein receptor; angiogenesis; angiostatin;  
KM endostatin; platinogen, laminin, treatment; wound healing; solid tumour  
KM psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;  
KM cerebral collateral; arteriovenous malformation; rheiosis; cancer;  
KM diabetic retinopathy; arthritis; wound healing; peptic ulcer;  
KM Helicobacter related disease; fracture; cat scratch fever.

OS Homo sapiens.

PN WO200032631-A2

PD 08-JUN-2000

PF 06-DEC-1999; 99WO-US28897.

PR 04-DEC-1998; 98US-0206059.

PA (ENTR-) ENTREMED INC.

PI Macdonald NJ, Sim KL;

DR WPI; 2000-412290/35.

PT New angiogenesis-inhibiting protein receptors useful in methods for  
 PT treating diseases and processes that are mediated by angiogenesis, such  
 PT as solid tumours, psoriasis, scleroderma and myocardial angiogenesis -  
 XX  
 PS Disclosure; Figure 3; 100pp; English.

CC This invention relates to angiogenesis-inhibiting protein receptors, and  
CC the DNA sequences encoding them. Angiogenesis is the generation of new  
CC blood vessels into a tissue, and normally occurs in wound healing,  
CC foetal and embryonal development and the formation of the corpus luteum,

CC endometrium and placenta. Angiostatin is a protein (see ABB16450 and  
CC A6689202) involved in angiogenesis, and has an amino acid sequence  
CC similar to that of a plasminogen fragment (see murine plasminogen  
CC AAB16490). Angiostatin has the ability to inhibit angiogenesis.  
CC Endostatin is also an angiogenesis inhibiting protein (see ABB16451 and  
CC A6689203). Sequences A668242 and AAB16522 represent coding and protein  
CC sequences of human laminin. Laminin is an angiostatin binding protein,  
CC and some of the peptides of the invention share homology with regions of  
CC laminin. Peptides ABB16452-B16521 (excluding ABB16490) are the  
CC angiogenesis-inhibiting protein receptor fragments of the invention. The  
CC peptides bind either angiostatin or endostatin and can be used in methods  
CC for treating diseases and processes that are mediated by angiogenesis,  
CC such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,  
CC Crohn's disease, cerebral collateral, arteriovenous malformations,  
CC rheusitis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,  
CC Helicobacter related diseases, fractures, placentaion and cat scratch  
CC fever. They are useful for the detection and prognosis of cancer. DNA  
CC sequences A628204-A628241 encode the peptides of the invention.

**SQ Sequence 183 AA**

Query Match	100.0%	Score 346;	DB 21;	Length 183;
Best Local Similarity	100.0%	Pred. No. 1.2e-38;		
Matches	63;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

QY 1 ADRAAVPIVNLKDELLFSSSEGLKPGARI FSPDGKDVLRHPTWPKSWHGS 60  
64 ADRAAVPIVNLKDELLFSSSEGLKPGARI FSPDGKDVLRHPTWPKSWHGS 122

QY	61	DPN	63
	111		
Db	124	DPN	126

Search completed: September 10, 2003, 14:12:30  
Job time : 42 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)